Chemoenzymatic Synthesis of Azaphilone Natural Products

by

Joshua B. Pyser

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2021

Doctoral Committee:

Professor Alison R. H. Narayan, Chair
Professor Anna Mapp
Professor John Montgomery
Professor Pavel Nagorny

Joshua B. Pyser

pyserjb@umich.edu

ORCID iD: 0000-0003-2884-5112

To my mother, Michelle, my father, Neil, my brother, Sam, and my cat, Finn.

Thank you for supporting me on this journey.

Acknowledgements

This work would not have been possible if it were not for the love and support of so many people over the course of my life, but particularly since I began my time at the University of Michigan. Firstly, I would like to thank my advisor, Professor Alison Narayan, for her guidance and support. When I first joined Narayan Lab, I didn't even know how to run a column, but she took a chance on me anyway and I will be forever grateful for the opportunity to work in her lab. I have learned so much from her over the last five years, and I am incredibly appreciative of the hard work she has put in to build the lab into what it is today. I would also like to thank Professor David Sherman and Professor Anna Mapp for opening their labs to me for a semester long rotation: it was a great experience to work with them and their students and gain useful skills in areas I might not have otherwise.

I owe thanks to Professor Mario Wriedt, who welcomed me into his lab as an inexperienced undergraduate and allowed me to continue to perform research over a summer session. It was in his lab that I truly got a taste of chemistry research and saw what the world of modern science really looked like. Though perhaps I was not the most dedicated student at the time, I'm appreciative of the opportunities he presented me with to better myself and advance my career. I would also like to thank my high school chemistry teacher, Mr. Shawn Therrien, for sparking my interest in chemistry in the first place. He was an excellent teacher who cared deeply about his students, made learning fascinating and fun, and was a "beast" at grading our tests.

Being in the second class of students in Narayan Lab, I am thankful for the members of the first class (Dr. Tyler Doyon, Dr. April Lukowski, Dr. Stephanie Chun, and Dr. Summer Baker Dockrey) for paving the way and helping to start some excellent projects. I am grateful for all you have taught me and being there for me when I was unsure or needed help. As for the rest of the lab, both new members and old ones, it's been a joy to have worked with you over the last five years. Thank you for making the lab a fun and comfortable place to be, you are all like my family now and I will certainly miss you (particularly the PM shift!). Dr. Suman Chakrabarty and Evan Romero, thank you for your contributions to both the figures and text of the first chapter of this work, it was a great experience to write this piece with you. I would also like to thank Dr. Ye Wang for his contributions to the fourth chapter, I've learned so much in the short amount of time working with him and we've made a great team making molecules together.

I overcame many challenges over the course of this work, but there were some I simply could not have without the help of my collaborators. I would like to thank Dr. Leo Joyce and Dr. Ren Wiscons for helping solve some absolute configuration conundrums, as well as Jesse Wotring for agreeing to perform bioassays for us and Dr. Troy Wymore for indulging my curiosity about computational modeling. Dr. Wendy Feng taught me an incredible amount in regard to MS and NMR analysis, and I would like to thank her for being ready at nearly a moment's notice to fix whatever problem I was having with an instrument.

Finally, I would not be here writing this if it were not for my family. My mom, dad, and brother have been my rocks throughout this entire experience, I am so grateful for all you have done to help make my life outside the lab easier and more fun. It was such a

relief to have a freezer full of home cooked food to come home to after a long day or extra pairs of hands to help me move, but those things pale in comparison to the emotional support, encouragement, and love you've provided over the years. You believed in me even when I didn't believe in myself, and I could not have asked for better parents, friends, and role models. Thank you.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AT – acyltransferase

AU – absorbance unit

DAD – diode array detection

E. coli – *Escherichia coli*

FDMO – flavin-dependent monooxygenase

HCl - hydrochloric acid

HTS – high-throughput screening

IPTG – isopropyl β- d-1-thiogalactopyranoside

Kan – Kanamycin

LB – Luria Broth

LC – liquid chromatography

MS – mass spectrometry/mass spectrometer

NADP+/NADPH – nicotinamide adenine dinucleotide phosphate/nicotinamide adeninedinucleotide phosphate

NMR – nuclear magnetic resonance spectroscopy

PigD – MrPigD

psi – pounds per square inch

SAR – structure-activity relationship

SSN – sequence similarity network

UPLC – ultra-performance liquid chromatography

µg – microgram

xg – g force

Abstract

Over the course of evolution, Nature has crafted innumerous enzymatic machines, capable of great feats of chemistry, across nearly all walks of life. These biocatalysts can construct incredibly complex molecular scaffolds, as demonstrated by the abundance of intricate natural products isolated from producing organisms. These compounds and the molecular machinery responsible for building them have inspired synthetic chemists for decades, but it is only more recently that technology has advanced to the point where the use of enzymes for organic synthesis has become feasible. Not only do biocatalysts often offer advantages over their small molecule counterparts in terms of ease of use, more mild conditions, greater selectivity, and faster reaction times, but also in the sense that many have evolved to operate synergistically with one another. This inherent coaction opens the possibility of performing enzymatic and/or chemoenzymatic cascades to construct complicated molecules in a one-pot fashion, chaining together multiple transformations and potentially streamlining chemical syntheses. Chapter 1 summarizes the history of biocatalysis in the context of organic synthesis, the current state of this field, and future prospects for this methodology in the synthesis of complex molecules.

The remainder of this thesis describes the use of enzymes from several biosynthetic pathways to construct members of an underexplored class of fungal natural products called azaphilones. This large family of compounds contains hundreds of scaffolds with unique structural features that impart a wide array of biological properties.

Though azaphilones represent a potentially untapped pharmacophore, their construction has proven challenging through conventional chemical syntheses.

In Chapter 2, I describe our rationale behind the motivation for using biocatalysis to construct azaphilones, the initial challenges associated with this goal, and our search for enzymatic catalysts to mediate the necessary dearomatization reaction towards constructing the azaphilone core. Using a sequence similarity network (SSN), we identified several previously uncharacterized flavin-dependent monooxygenase (FDMO) enzymes capable of mediating the desired reaction in a model system. Further analysis of these FDMOs also provided evidence for structural features of these enzymes that control the facial selectivity of the dearomatization reaction.

Having identified and characterized useful FDMO enzymes, we then constructed several azaphilone natural products through chemoenzymatic syntheses in Chapter 3. Two FDMO homologs, AzaH and AfoD, provided access to both enantiomers of an azaphilone bicycle, upon which the natural products trichoflectin and lunatoic acid A were built. Obtaining both enantiomers of trichoflectin through this methodology provided evidence for the structural revision of this compound, uncovering an error in the azaphilone literature on the absolute configuration assignment of these compounds. AzaH also provided access to another natural product, deflectin-1a, allowing for its structural revision, as well.

Finally, in Chapter 4, I describe the development of a dual-enzyme platform to construct azaphilone tricycle analogs. Several acyltransferases were identified using an SSN in an attempt to identify one suitable for the needs of this project. Ultimately, the acyltransferase (AT) MrPigD (PigD) was investigated for its ability to acylate the C7

hydroxyl group of azaphilone bicycles. Our studies indicate that this enzyme operates synergistically with AzaH in a one-pot sequential reaction, providing acylated azaphilones directly from five different orcinaldehyde precursors. This sequence platform also provided access to the natural product, rubropunctatin, through an unexpected cyclization of the acylated intermediate to afford a linear tricycle. Following analysis of the compound by UPLC, quantities of the natural product sufficient for NMR studies were obtained through a preparative scale biocatalytic reaction. This method was also applied to the synthesis of two other tricycle analogs, whose ^{1}H NMR spectra suggest they are also linear tricycles.

The work presented in this thesis demonstrates the utility of leveraging FDMO and AT enzymes to construct complex azaphilones with greater ease and selectivity than achieved through conventional chemical syntheses. We anticipate that these methods will provide access to diverse libraries of molecules in a manner that is orthogonal to high-throughput screening (HTS) towards in-depth structure-activity relationship (SAR) studies.

Chapter 1: Introduction

With excerpts from “State-of-the-art Biocatalysis”. Pyser, J. B.; Chakrabarty, S.; Romero, E. O.; Narayan, A. R. H. *ACS Cent. Sci.* **2021**, *Manuscript accepted.*

Summary

Enzymes have been used by humanity since our inception, and as our understanding of these macromolecular machines develops, they will continue to grow in their value to organic chemists. In addition, with the growing need to develop green alternatives in synthetic chemistry, biocataylsis represents a promising avenue to achieve this goal without sacrificing the activity and selectivity of conventional small-molecule catalysts. From its humble beginnings affecting simple manipulations like kinetic resolutions or reductive aminations in an academic setting, biocatalysis has now started to expand to the broader chemistry community. The internet and globalization have made it simple for anyone to obtain the tools and knowledge needed to use enzymes, greatly expanding the interest in and study of this field. State of the art techniques in chemoenzymatic synthesis and enzymatic cascades have now been applied in industrial processes to perform transformations with impressive yields and efficiency *en route* to highly valued drug scaffolds. As the technology continues to evolve, biocatalysis will provide new opportunities in high-throughput drug synthesis and screening, as well help

to streamline complex molecule syntheses, reinforcing the need to support the growth of these next-generation methodologies.

1.1 Background

The utility of naturally-occurring enzymes has been harnessed for thousands of years through fermentation and food preservation processes.[1] Fascination with the chemistry of microbes originated at the dawn of the Neolithic era, nearly 12,000 years ago, when humans began domesticating grains and consuming alcohol, the evidence of which can be found in archaeological records.[2] In fact, arguments have been made that our use of alcohol produced enzymatically even predates archaeological records, with evidence for ethanol-degrading enzymes present in primate species that lived before *Homo sapiens*.[2] Driven by curiosity, people sought to understand how and why leaving cereals or grapes[3] alone for some time caused them to adopt new properties, seeding fields like enzymology, molecular biology, and biocatalysis as an extension of this fascination.[4]

Within the past few decades, biocatalysis in fine chemical and pharmaceutical production has surged.[5-7] This trend is driven in part by advances in DNA sequencing, bioinformatics and protein engineering that allow for the identification of enzymes that

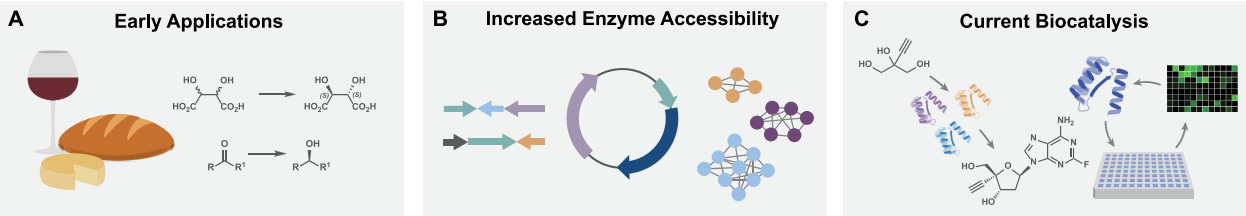


Figure 1.1. A. Early uses of enzyme-mediated transformations, such as fermentation, chiral resolutions, and functional group interconversions. B. Recent advances in genome sequencing, gene synthesis, and bioinformatics increase the accessibility of obtaining enzymes. C. Select strategies in modern biocatalysis include cascades, chemoenzymatic synthesis, and enzyme evolution.

meet the reactivity and selectivity needs of a given synthetic route.[8] Biocatalytic reactions are now routinely used in scalable processes ranging from simple chemical manipulations such as chiral resolutions,[9-11] reductive aminations,[9, 12] and alcohol oxidations,[13] to complex, multistep chemoenzymatic cascades that enable access to high-value drug molecules on an industrial scale.[14] The rapid developments in biocatalysis are also enabling a re-emergence of natural products in the current era of genomics.[15] Natural products have a long history in drug discovery and development given their often potent biological activity. Still, the structural complexity of these compounds challenges chemists and demands a substantial time and resource investment to synthesize these compounds and analogs thereof.[16] It seems undeniable that the next logical step in synthetic chemistry is to leverage the machinery that Nature has developed to access a similar breadth of complexity exhibited by these compounds.

Biocatalytic methods also offer several key advantages over traditional chemical processes. These advantages include increased safety and sustainability, procedural simplicity, and the tunability of an enzyme's reactivity or selectivity through protein engineering.[7] In particular, the sustainability profile of enzymatic transformations has motivated their adoption.[17] In contrast to the organometallic catalysts developed using precious metals, which are being rapidly depleted from the Earth's crust,[18] enzymatic catalysts can be produced without the worry of exhausting limited resources. These advantages poise biocatalysis for adoption into the mainstream synthetic repertoire.[19] It is easier now than ever before for anyone, from enzyme novices to global biotech companies, to tap into the powerful transformations biocatalysis can offer.

In Chapter 1, we explore the diverse field of biocatalysis and highlight recent advances in technology that have dramatically improved the accessibility of enzymes and driven the transition from simple biocatalytic systems to sophisticated complexity-generating biocatalytic platforms. Additionally, we highlight the advantages that biocatalysts offer in organic synthesis, the current state-of-the-art in this field, and how advancing technologies will provide new opportunities for incorporating biocataytic strategies in the synthesis of target molecules.

1.2 Early Applications of Biocatalysis in Synthesis

Despite their use in the fermentation process for millennia, enzymatic methods were first appreciated on the molecular level beginning in the early 1800s[20] when researchers began investigating yeast for their fermentation abilities.[21] Though it was first believed that the entire microorganism itself was functioning as the catalyst, the discovery of the first enzyme mixture, called "diastase", fundamentally changed the field by demonstrating that observed reactions were mediated by only specific parts of the organism.[22] This sparked increased interest in the then-new field of enzymology, and crucial milestones in understanding enzymes followed. These include the development of the lock and key model,[23] cell-free fermentation,[24] the realization that enzymes were in fact proteins,[25] and the elucidation of the DNA structure,[26] all of which have

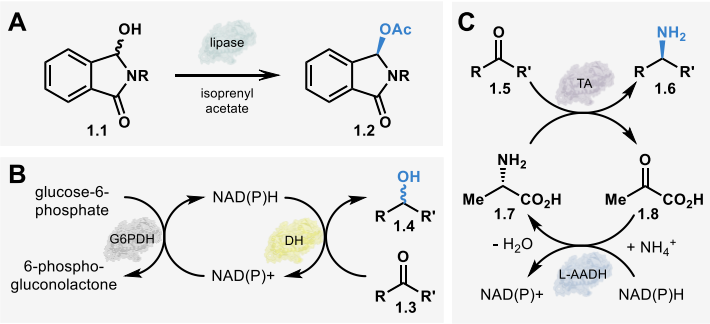


Figure 1.2. A. Dynamic kinetic resolution of racemic *N*-acylhemiaminals by a lipase. B. NAD(P)H recycling system developed by Wong and Whitesides. C. Cascade system for construction of chiral amines using an ω-transaminase. Abbreviations: G6PDH – glucose-6-phosphate dehydrogenase, DH – dehydrogenase, TA – transaminase, L-AADH – L-α-amino acid dehydrogenase

paved the way for modern biocatalysis.[21] Finally, with the invention and adoption of X-ray crystallography, researchers were finally able to view the three-dimensional structure of these miraculous macromolecules in detail and gain insight into their functions and mechanisms.[27] A more detailed account of the rich history of enzymology has been reviewed previously.[21]

Over much of the 20th century and into the early 2000's, the use of enzymes to perform useful chemistry truly gained popularity.[21] Enzyme-mediated kinetic resolutions were one of the most common initial uses of biocatalysis in synthesis. Though several different classes of enzymes have been applied to conduct these enantiomeric enrichments,[9] lipases are commonly employed to affect this transformation based on their commercial availability, large substrate scope, high levels of selectivity, and cofactor-free catalysis.[11] Also commonly used in the production of cheese products and laundry detergents,[28] the first member of this enzyme class was discovered in 1848 by Claude Bernard in his investigation of pancreatic secretions.[29] Initial experimentation with lipases in the 1930's[30] and 1940's[31] laid the groundwork for their use in kinetic resolutions and other biocatalytic transformations for the rest of the 1900's.[11, 28] A relatively recent example published by Kaga and co-workers in 2003 demonstrates the simplicity of using lipases in a more modern biocatalytic setting; to construct a small library of chiral hemiaminals through the dynamic kinetic resolution of racemic starting materials.[10] The group first screened a set of commercially available lipase enzymes for acylation activity against their library of racemic *N*-acylhemiaminals and determined that lipase QL gave short reaction times and operated with high levels of enantioselectivity. With this enzyme, they constructed several *O*-acylated hemiaminals in quantitative yields and in

high/exquisite enantioselectivities (Figure 1.2A). Lipase QL is just a select example among several others that demonstrates the early and widespread use of lipases and other kinetic resolution enzymes in academia and industry.[28]

The design of systems for *in situ* cofactor regeneration is a significant milestone in biocatalytic method development.[32] Early studies of cofactor-dependent enzymes in synthesis relied on the addition of stoichiometric quantities of these cofactors, which limited the utility of the enzymatic reaction. Thus, the ability to continuously recycle these essential components in the reaction mixture was critical to certain biocatalysts' practical use.[33] The early work of Wong and Whitesides on the regeneration of NAD(P)H *in situ* to enable reductions by dehydrogenase enzymes demonstrates this method's capability and has since made an enormous impact on the field.[34-36] To apply these dehydrogenases towards the construction of chiral alcohols, they developed the use of glucose-6-phosphate dehydrogenase (G6PDH) from *L. mesenteroides* to reduce the NAD(P)+ cofactor *in situ* following its oxidation by the dehydrogenase. G6PDH relies on glucose-6-phosphate, which is inexpensive and easy to synthesize, to provide the equivalent of hydride needed to reduce $NAD(P)^+$ to NAD(P)H (Figure 1.2B). With this recycling system in place, Wong and Whitesides completed the biocatalytic generation of optically pure D-lactic acid, *threo*-D_s(+)-isocritic acid, and (*S*)-benzyl-α-d_1 alcohol.[34] This early example of a biocatalytic cascade has since enabled the use of many enzyme-catalyzed reductions and has paved the way for application on industrial scale.[37]

Since this preliminary work, methods relying on electrochemistry, photochemistry, other hydride donor/acceptor systems have been developed. For example, readily available reagents like isopropanol have been used in cofactor regeneration systems,

providing an alternative to the more expensive and labor-intensive sugars used previously.[32] A variety of more economical and industrially-feasible sacrificial functional group donors have also been applied to improve efficiency, scalability, and ease of use of cofactor regeneration.[38, 39] Several reports describe the use of isopropyl amine as an amino donor for transamination reactions, providing a substitute for the cost prohibitive amino acids used conventionally.[40, 41] There is also a focus on the construction and regeneration of synthetic, biomimetic cofactors, which holds promise for increasing the effectiveness of these systems further.[42]

The rapid adoption of recycling system methods allowed for the broad application of biocatalytic reduction reactions.[37] The use of transaminases for constructing chiral amines is a prime example, and their utility in synthesis has been showcased in the synthesis of drug molecules such as sitagliptin.[43, 44] Transaminases offer many advantages over their chemical counterparts, including improved stereoselectivity, mild reaction conditions, and reducing the reliance on harmful solvents and transition metals.[43, 45] A notable example of the application of ω-transaminases, a subgroup of transaminases that has drawn particular attention in the pharmaceutical industry,[46] was developed by Koszelewski and co-workers. Ultimately relying on a system similar to that produced by Wong and Whitesides to power catalysis, the Koszelewski group constructed nine chiral amines with excellent enantiomeric excess through a reductive amination with the commercial ω-transaminase ATA-113.[12] They also employed a second enzyme, L-α-amino acid dehydrogenase (L-AADH), to further streamline this reaction by regenerating the amino acid alanine *in situ*, which is the amine source for the transamination reaction. L-AADH, enabled by the NAD(P)H recycling system, utilizes an equivalent of ammonium

as the ultimate nitrogen source to reduce pyruvate to the desired alanine (Figure 2C). Albeit a relatively simple transformation by today's standards, this early work serves as a quintessential example of synthetic utility of transaminases.

The seminal work highlighted here, alongside other early examples of simple biocatalytic reactions such as isomerizations, redox manipulations, and ligations,[47] brought to light the power of enzymes as catalysts in synthesis. In the more modern history of biocatalysis, there has been a paradigm shift from using enzymes to construct relatively simple building blocks or provide chiral intermediates for traditional syntheses,[48] to relying on them for late-stage synthetic modifications,[49] combining molecule fragments towards value-added compounds, and conducting multistep, biocatalytically-mediated total syntheses.[14, 37] Additionally, the tools for investigating and leveraging biocatalysts for synthetic uses have reached a stage where they are widely accessible to the chemistry community: obtaining the knowledge and equipment needed for biocatalysis can be accomplished with just a few clicks.

1.3 Accessibility of Biocatalysis to Synthetic Chemists

Once relegated to the fields of biochemistry and molecular biology, recent advances in bioinformatics,[50] DNA sequencing,[51] protein engineering,[52] and DNA synthesis have made it possible for virtually anyone to take advantage of enzymatic catalysts and tailor them to their own needs. The process of identifying, producing, isolating, and tuning the reactivity of biocatalysts for desired transformations is as accessible to synthetic chemists as obtaining and using small molecule catalysts. In particular, the recent exponential growth in annotated protein sequences available in

online databases has created an enormous catalog of potential enzymes to serve many synthetic needs. Two of the most popular databases, UnitProt[53] and Genbank[54], now house information on more than 420,000 individual species, representing over one billion total sequence records. Instead of taking to the field and collecting specimens by hand to examine their genes, these databases store a wealth of information on protein sequence and origin and are a valuable starting point for anyone looking to identify enzymes for a given synthetic purpose.[55]

Combining the vast amount of data stored in these online libraries with bioinformatic tools allows one to begin making predictions about the function of uncharacterized or "hypothetical" proteins,[56] and to search for previously identified proteins that may also demonstrate activity in a non-canonical transformation.[57] For example, the basic local alignment search tool (BLAST) is one of the most popular and easy to use for this type of analysis.[58, 59] Gaining popularity in the early 1990s and now available to use for free on the National Center for Biotechnology Information (NCBI) website[59], this tool relies on algorithms to search available online databases for protein sequences that resemble a given input sequence. By feeding the BLAST search engine a known nucleotide or amino acid sequence, or a protein identifier such as an accession number, the tool can align all known protein sequences that share similarity with the input sequence and rank them in a list. As minute changes in the order or position of amino acid residues can drastically alter function between homologous proteins with highly similar sequences, this type of search can be advantageous when trying to identify enzymes with improved stability and activity, complementary substrate scopes, or proteins that can perform desired transformations with the alternative site- and/or

stereoselectivity to the one used to build the query.[36, 60, 61] This tool also provides known information about each sequence, such as the originating organism and any characterized metabolic function of the protein within said organism. By displaying data on the degree of similarity between proteins based on how well their sequences align, a user can quickly identify any known proteins that may share functional characteristics with the input protein sequence.[62]

Albeit a useful starting point, this list format provided by BLAST can become cumbersome when the search yields thousands of potentially related protein sequences. To obtain a more comprehensive view of entire protein families, some of which can contain hundreds of thousands of proteins,[63] tools have been developed that provide greater context for viewing connections within these groups. Phylogenetic trees are commonly used to examine relationships between homologous proteins and study

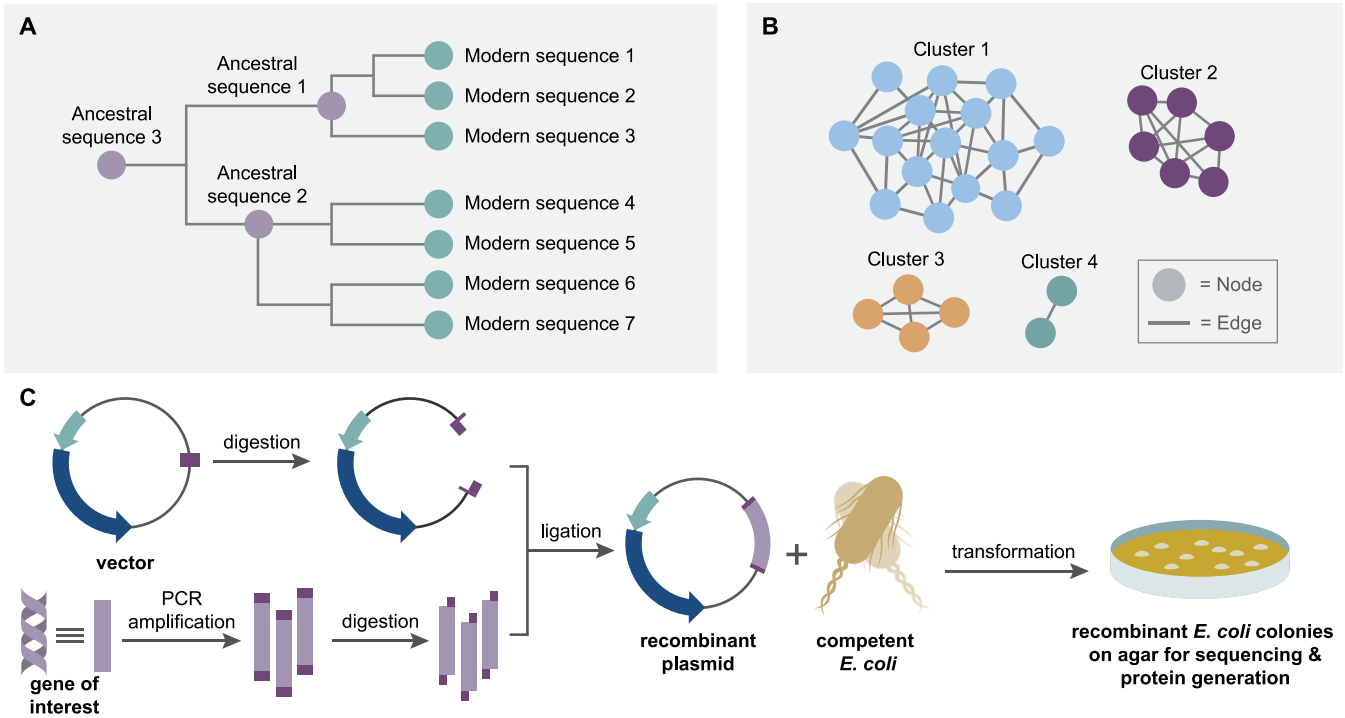


Figure 1.3. A. Conceptual phylogenetic tree depicting locations of calculated ancestral sequences. B. Conceptual SSN demonstrating nodes, edges, and clusters. C. Workflow for a traditional cloning procedure.

changes in protein families over their evolution.[64] This bioinformatic analysis technique relies on the alignment of homologous protein sequences to construct a visual representation of the evolutionary history of the related sequences in a phylogenetic tree (Figure 3A).[64] Building and visualizing these trees has also been simplified by programs like the Molecular Evolutionary Genetics Analysis (MEGA)[65, 66] and Ensembl[67] that provide straightforward user interfaces. Once various algorithms and search tools are applied to analyze all available data and establish the most likely configuration, the trees can be examined to draw conclusions about relatedness among protein families and test hypotheses about their evolutionary origins.[68, 69]

For example, one intriguing use of phylogenetic analyses in the context of biocatalysis is the identification and reconstruction of ancestral protein sequences (Figure 1.3A) that can offer benefits in stability and biocatalytic activity over their modern "offspring".[70] This technique relies on software to compare related protein sequences that most likely evolved from a common ancestor to calculate or "infer" the exact sequence of that ancestral protein.[71] The ability to now obtain any DNA sequence quickly and easily makes reconstructing ancestral proteins a potentially powerful tool in identifying novel enzymes with desirable functions. To this effect, Furukawa *et al.* have identified an ancestor of 3-isopropylmalate dehydrogenase (IPMDH), a key enzyme in the biosynthesis of leucine, which offers improvements in its stability and activity over extant IPMDH enzymes from present-day organisms through construction and analysis of a phylogenetic tree.[72] Following inference and identification of two ancestral protein sequences, dubbed ancIPMDH-IQ and ancIPMDH-ML, the group successfully expressed each protein in *E. coli* and, after isolating the enzymes for further investigation, discovered

they provided increased thermal stability and improved catalytic activity at low temperatures compared to their modern homologs.[72] This work demonstrates just one of many potential uses for ancestral protein reconstruction, as other reports describe how ancestral proteins might possess higher degrees of substrate promiscuity compared to their modern offspring, thus offering potentially valuable characteristics to organic chemists seeking diverse and novel bond-forming activity.[73]

Despite their utility and newfound ease-of-use, phylogenetic trees can still prove overwhelming when examining extensive protein families or groups of sequences.[74, 75] Tools like sequence similarity networks (SSNs) have emerged to help overcome these challenges. SSNs have gained much attention since its introduction to the bioinformatics community in 2003.[76] It provides a way to visualize family-wide relationships and patterns in large groups of protein sequences by ranking sequences in "clusters" based on their alignment scores.[74-77] These networks comprise groups of "nodes," representing a protein sequence or group of sequences. These nodes are then connected by lines called "edges", representing a threshold for sequence similarity that can be set by the user (Figure 1.3B). Changing this score controls which nodes group together, allowing for inferences to be made about protein structure and functions by examining and comparing the location of nodes within the clusters.[77] These networks can be constructed and analyzed quickly and easily through a web-based tool called EFI-EST[75] and the free-to-download software Cytoscape.[76] Helpful tutorials and videos on how to construct, use, and manipulate SSNs with these programs are also available for free online.[75, 76]

These networks can be beneficial for chemists looking to identify new enzymes for catalysis from families with a limited number of previously characterized proteins. Lewis

and co-workers have recently applied SSNs to identify and profile novel flavin-dependent halogenase (FDH) enzymes.[78] Using these networks to guide their search, the group elected 128 initial halogenase sequences to sample for useful halogenation activity. Following expression of the genes, they obtained 87 soluble proteins for preliminary activity screens with 12 initial substrates containing a mixture of phenols, indoles, and anilines. Overall, the group identified 39 previously uncharacterized halogenases that demonstrated unique bromination and/or chlorination activity against the substrate panel. After examining an additional 50 complex and bulky substrates, they discovered at least one member of their halogenase library that demonstrated activity with around 48% of the substrates tested. Ultimately, Lewis and co-workers examined and characterized the preference for these FDHs towards bromination and chlorination, their site-selectivity, and thermostability and could draw further conclusions about trends in their SSNs through this family-wide profiling.[78] This cutting-edge application of SSNs demonstrates how free and straightforward internet-based software can be used to identify synthetically tractable biocatalysts without the need to perform more complex mutagenesis and directed evolution experiments.

Our group has also demonstrated the applicability of SSNs to examine previously uncharacterized enzymes with useful chemical functions.[36, 74] We sought to identify homologous flavin-dependent monooxygenase (FDMO) proteins to investigate the factors that control their site- and facial selectivity in an oxidative dearomatization reaction and to identify enzymes suitable to enable a stereodivergent chemoenzymatic natural product synthesis campaign.[36] Analysis of an SSN comprised of over 45,000 sequences from the flavin adenine dinucleotide (FAD) binding domain protein family (pfam01494)

identified several FDMOs that are highly similar to those our group had investigated previously.[35] Combining the experimental data gained from reactions of these enzymes in a model system with comparisons of their sequence information and location in the SSN allowed us to identify trends in the SSN that predict the site-selectivity of a putative FDMO based on which cluster it is located in. We envisioned this technique may also help predict the stereoselectivity of the dearomatization mediated by a given FDMO, but further studies suggest that this is much more finely controlled than what can be predicted by a precursory SSN. Additional studies suggested two key active site residues are crucial in controlling the stereochemical outcome of the dearomatization reaction known to these proteins.[36, 60] Though this does highlight a potential drawback of using SSNs in this way, the tool did ultimately demonstrate its utility in identifying other catalytically active proteins with desired activity. Work is currently underway to further characterize these enzymes in hopes of expanding our library of biocatalysts.

Before developing these tools for identifying and characterizing enzymes *in silico*, obtaining biocatalysts for chemical experimentation was a significant challenge. To investigate a wild-type or naturally occurring catalytic protein, a molecular biologist would first need to get the source DNA or RNA encoding the gene of interest from the native organism. Following isolation, it is necessary to amplify the DNA fragment through a polymerase chain reaction (PCR)[79]. These amplified fragments must then be digested with restriction enzymes and ligated into a circular piece of DNA called a plasmid that has also been prepared with the same enzymes to ensure the ends of these sequences are compatible.[80] Inserting DNA into a vector such as this not only allows for the host organism to uptake the gene of interest but can also be used to impart properties like

antibiotic resistance to the transfected cells to allow for the selection of individual cells that have successfully incorporated the plasmid. Following digestion, the prepared DNA fragment and cut vector are then combined in the presence of a DNA ligase enzyme, which efficiently joins the compatible ends of the fragment and vector, resulting in the production of a so-called "recombinant plasmid".[81] In the case of transforming *E. coli*, one of the most popular and easy-to-use host organisms for recombinant protein production, the recombinant plasmid is then added to competent bacterial cells (cells that are primed to uptake foreign DNA from their surroundings). The cells can then be grown on agar media possessing an antibiotic to prevent cells that do not contain the plasmid from growing. After allowing the cells to grow on the agar, a colony can be harvested and analyzed to ensure that it possesses the desired gene. Finally, after ensuring the gene is present and contains the correct sequence, the colony can be used to seed a larger culture to harvest usable amounts of the desired protein, as well as to produce more of the plasmid for additional studies or to transfect new cells with the desired gene without having to undergo the entire process from scratch (Figure 1.3C).[81]

In contrast to these traditional cloning techniques, technological breakthroughs in modern gene synthesis provide a highly streamlined process for chemists seeking DNA sequences and plasmids. Instead of using isolated DNA from native organisms as a template to amplify, solid-phase oligonucleotide synthesis allows for the *de novo* construction of any nucleotide sequence found online from individual nucleotide bases.[82] Companies now offer customized DNA constructs for purchase on-demand: input your insert sequence of interest and choose the desired vector in their online interface, and the company will ship you a ready-to-use recombinant plasmid possessing your exact

gene, or even a sample of host organisms containing the plasmid, in a matter of weeks. The cost is dependent on the number of base pairs in the DNA sequence and the particular plasmid desired, but these DNA constructs can typically be purchased for under 200 USD. Not only does this save time and effort in obtaining the recombinant vector, it also allows for nearly anyone to take advantage of this technology without the need for specialized equipment, reagents and knowledge required for traditional cloning. Inexpensive and straightforward methods, reagents, and equipment for transforming, growing, and isolating recombinant protein from cells containing a mail order plasmid also lower the barrier for individuals and labs looking to enter the field of biocatalysis.[83, 84]

These tools and techniques described above barely scratch the surface of what is available for anyone interested in using and tuning biocatalysts for a particular synthetic application.[75, 85, 86] Advances in the fields of directed evolution[52] and computer-guided enzyme engineering[87] promise to construct enzymes with ever-greater efficiency, selectivity, stability, and reusability than those known today. Leveraging combinations of these strategies have already begun to provide highly applicable and useful biocatalysts to the synthetic community at large and will continue to improve biocatalytic methods as they are developed further.

1.4 State of the Art Biocataklysis

Following this explosion of interest in enzyme-mediated catalysis, biocatalytic reactions are now increasingly employed in complex molecule synthesis. Biocatalytic methods that affect late-stage site- and stereoselective C–H functionalization constitute one of the best state-of-the-art transformations available today that maximize step

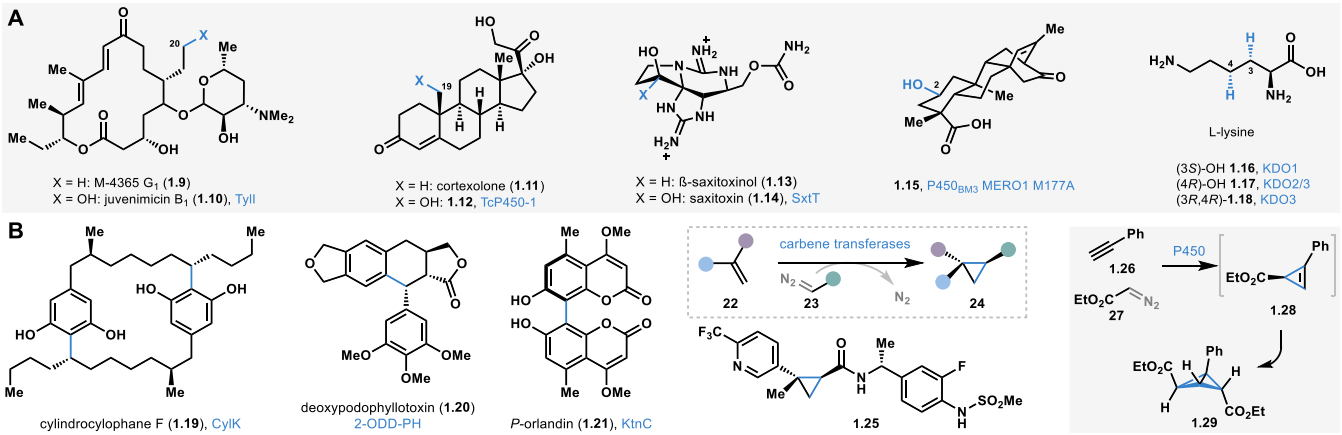


Figure 1.4. Biocatalysis in complex molecule synthesis: A) Selected C-H functionalization reactions. B) Selected C-C bond forming reactions.

efficiency and enable diversification of complex scaffolds. Select examples of biocatalytic C–H functionalization in complex molecule synthesis are shown in Figure 1.4A. Sherman and co-workers have carried out a late-stage hydroxylation of the macrolide natural product M-4365 G1 (**1.9**) to form antibiotic juvenimicin B1 (**1.10**) with P450 monooxygenase TylI.[88] Late-stage biocatalytic C–H hydroxylation has also been explored in the pursuit of steroid-based drugs.[89, 90] Zhou and co-workers developed a biocatalytic C19 hydroxylation of cortexolone (**1.11**) to form 19-hydroxycortexolone (**1.12**) using TcP450-1, a cytochrome P450 enzyme.[90] This strategy enables direct access to bioactive C19-hydroxylated steroids.[90] It is worth mentioning that direct hydroxylation at the C19 position of steroids is extremely challenging using traditional chemical methods.[91-93] Our research group's long-standing interest in using enzymes to carry out C–H hydroxylation reactions has been channeled for the late-stage diversification of paralytic shellfish toxins.[94-97] We have employed the Rieske oxygenase SxtT to carry out the site- and stereoselective hydroxylation of β-saxitoxinol (**1.13**), directly generating saxitoxin (**1.14**).[95] The Renata group recently disclosed a non-heme iron (NHI) dependent enzymatic platform to enable late-stage biocatalytic hydroxylation of complex terpene scaffolds.[98]

The enzyme $P450_{BM3}$ MERO1 M177A was employed in carrying out selective C–H hydroxylation to form the oxidized terpene product **1.15**.[98] Direct C–H hydroxylation has also been developed for amino acid scaffolds. For example, Zaparucha discovered the NHI enzyme KDO1-3 that carried out selective hydroxylation of L-lysine.[99, 100] The enzymes KDO1 and KDO2/3 selectively hydroxylated the C3 and C4 positions of L-lysine, and the enzyme KDO3 carried out C4 hydroxylation of a pre-C3-hydroxylated L-lysine.[99] Renata and co-workers employed the KDO1 mediated C3-selective hydroxylation of L-lysine in their total synthesis of tambromycin.[101]

Rapid advances in biocatalysis have resulted in the identification of enzymes that can carry out carbon-carbon (C–C) bond-forming reactions (select examples in Figure 1.4B).[102] Balskus and co-workers reported the enzyme CylK that carries out biocatalytic intermolecular Friedel-Crafts alkylation of two halogenated resorcinol derivatives to construct the cylindrocyclophane **1.19**.[103] The enzyme CylK has also been shown to be highly promiscuous, carrying out alkylation of a variety of resorcinol derivatives with secondary alkyl halides.[104] Biocatalytic Friedel-Crafts alkylation has also been carried out to synthesize podophyllotoxin lignans.[105, 106] For example, the NHI enzyme 2-ODD-PH has been utilized to carry out the biocatalytic synthesis of deoxypodophyllotoxin (**1.20**) and related analogs.[106-108] Biocatalytic oxidative phenolic coupling reactions are emerging as powerful tools to construct complex molecules.[109-111] The Müller group recently reported fungal P450 enzymes capable of carrying out oxidative coupling of coumarin derivatives in a site- and stereoselective manner.[109] For example, the enzyme KtnC catalyzes the synthesis of the bicoumarin *P*-orlandin (**1.21**).[109] Biocatalytic C–C bond formation has been explored in carbene transfers to generate chiral cyclopropanes.[112-114]

Arnold and co-workers first reported an engineered $P450_{BM3}$ that carried out carbene transfer reactions. Diazoacetate reagents were used as the carbene sources to carry out alkene cyclopropanation.[112] Several other groups have contributed to the development of biocatalytic carbene transfer reactions, and these have been applied towards the synthesis of pharmacologically relevant compounds such as the TRPV1 inhibitor **1.25**.[115, 116] Biocatalytic carbene transfer reactions can be extended to alkynes as well, where the first carbene transfer generates a cyclopropene product which is primed for a second carbene transfer reaction to generate stereopure bicyclobutane products.[117] This transformation rivals the best of what synthetic chemistry has to offer in terms of building complexity through C–C bond formation.

In the case of selective C–H functionalization and C–C bond-forming reactions, biocatalysis is often employed at an advanced stage or in the final step of a synthetic campaign. Alternatively, biocatalysis can be engaged at an early stage in chemoenzymatic synthesis planning (Figure 1.5). In such cases, the product of a biocatalytic reaction is transformed into a target molecule of interest using modern synthetic organic chemistry tools. This strategic merge of biocatalysis and small molecule-based synthetic methods enables access to chemical scaffolds previously

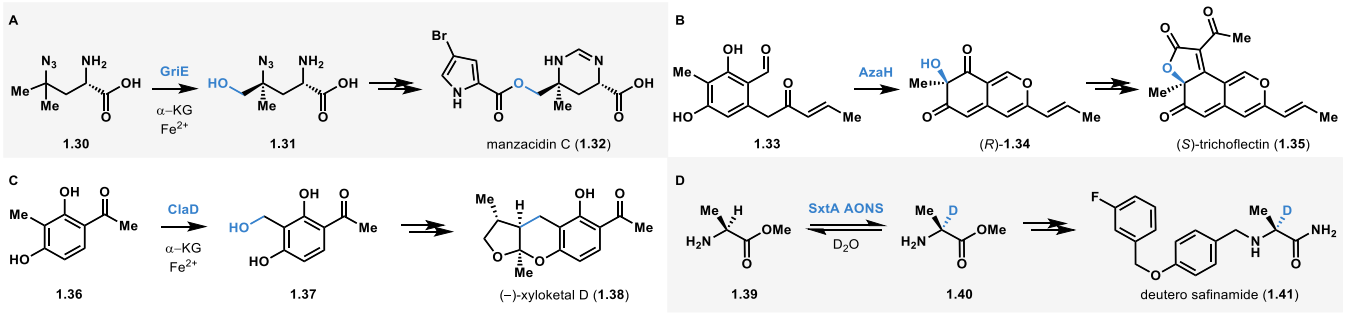


Figure 1.5. Chemoenzymatic sequences to complex molecules. A. Amino-acid C-H hydroxylation in the synthesis of manzacidin C. B. Hydroxylative dearomatization in the synthesis of azaphilone natural products. C. Benzylic hydroxylation en route to xyloketal D synthesis. D. Alpha deuteration of amino acids in the formation of deutero safinamide.

unattainable using traditional chemical methods alone. For example, Renata and co-workers developed a chemoenzymatic total synthesis of the natural product manzacidin C (**1.32**).[118, 119] The NHI-dependent enzyme GriE was employed to carry out selective hydroxylation of an L-leucine derivative **1.30** to form **1.31**.[118] The product **1.31** was taken through established synthetic steps to formally assemble manzacidin C (**1.32**).[118] Our group has been interested in the hydroxylative dearomatization of resorcinol compounds using flavin-dependent monooxygenases (FDMOs).[35, 36] We have employed the site- and stereoselectivity of FDMOs in conjunction with small-molecule-based methods to enable the total synthesis of azaphilone natural products.[36] For example, the enzyme AzaH was used to carry out the dearomatization of resorcinol **1.33** to form **1.34**. The quinol product **1.34** was subsequently transformed to (*S*)-trichoflectin (**1.35**) using chemical methods.[36] Our group has also focused on developing benzylic hydroxylation of *o*-cresol compounds using NHI-dependent monooxygenases.[120] For example, we have employed the enzyme ClaD to carry out benzylic hydroxylation of resorcinol derivative **1.36**, the product of which (**1.37**) undergoes spontaneous loss of water resulting in a biocatalytically generated *o*-quinone methide, which was trapped using a chiral dienophile to construct the bioactive natural product xyloketal D (**1.38**).[120] α-Deuterated amino acids are important building blocks towards the synthesis of labelled pharmaceuticals and biological probes, however, traditional methods to access these compounds often require protecting group manipulations[121] and can be difficult to perform in a stereoselective manner.[122] We discovered that SxtA AONS, α-oxoamine synthase evolved for saxitoxin biosynthesis, is capable of deuterating a range of unprotected amino acids and their methyl esters using D_2O as the deuteron source. For example, deuteration of alanine methyl ester (**39**)

resulted in **1.40**, which was subsequently transformed using chemical methods to access the deuterium-labeled Parkinson's pharmaceutical safinamide (**1.41**).

Multi-enzyme cascade reactions have been developed in industrial and academic laboratories to enable complex molecule synthesis (select examples in Figure 1.5). The process toward HIV treatment drug islatravir (**1.48**) developed by Merck and Codexis is a representative example of a multi-enzyme cascade employed on an industrial scale.[14] The artificial nucleoside islatravir (**1.48**) was constructed using a combination of five enzymes from the nucleoside salvage pathway in bacteria, which were each engineered for a distinct purpose.[14] This protecting group-free cascade yielded the product islatravir in markedly higher yields than previous chemical syntheses .[14, 123] Moore and co-workers developed a multi-enzyme synthesis of complex halogenated bacterial meroterpenoids napyradiomycins A1 and B1 (**1.54** and **1.55**) in a single pot.[124] Starting with three organic substrates (tetrahydroxynaphthalene **1.49**, dimethylallylpyrophosphate, and geranyl

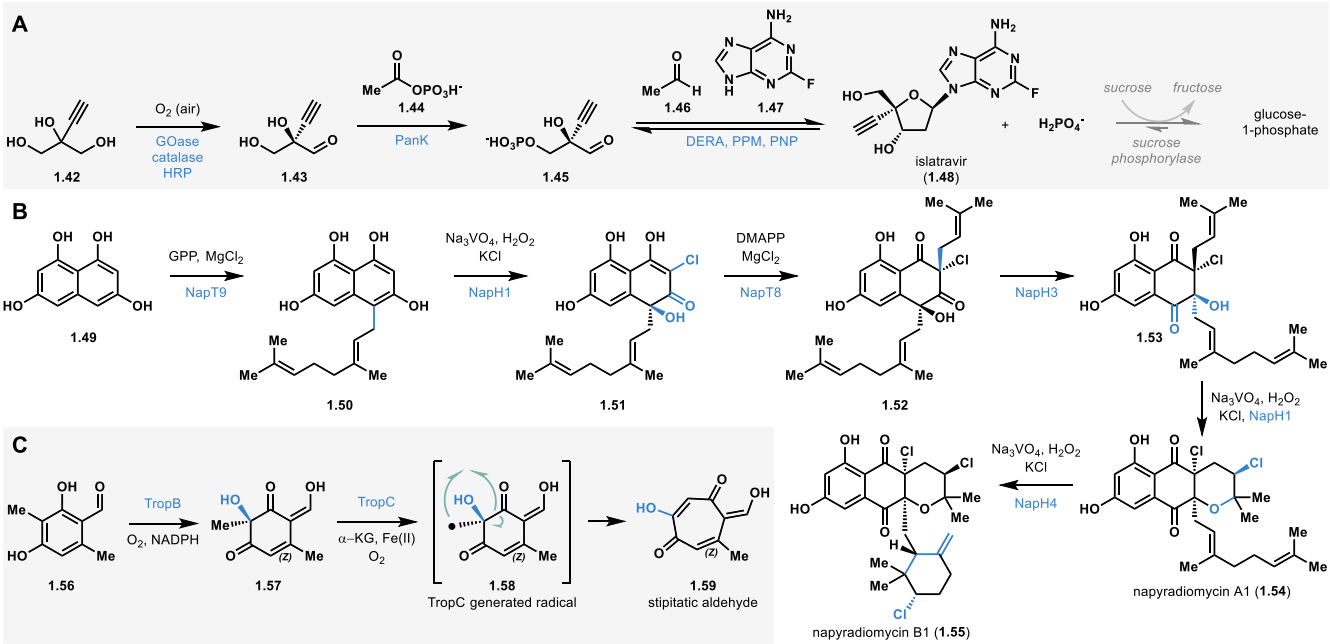


Figure 1.6. Multi-enzyme biocatalytic sequences: A. Merck's biocatalytic synthesis of islatravir. B. Multi-enzyme synthesis of napyradiomycin A1 and B1. C. Multi-enzyme sequence towards the synthesis of tropolone stipitatic aldehyde.

pyrophosphate), the team developed a catalytic sequence involving five enzymes: two aromatic prenyltransferases (NapT8 and T9) and three vanadium dependent haloperoxidase (VHPO) homologs (NapH1, H3, and H4) to assemble the complex halogenated metabolites in milligram quantities.[124] Our group has leveraged the exquisite reactivity of FDMOs and NHI-dependent monooxygenases to construct tropolone natural products.[35, 125] Tropolones are a structurally diverse class of bioactive molecules that are characterized by a cycloheptatriene core bearing an α-hydroxyketone functional group. We developed a two-step, biocatalytic cascade to the tropolone natural product stipitatic aldehyde starting with the resorcinol **1.56**. Hydroxylative dearomatization of **1.56** using TropB affords the quinol intermediate **1.57**. The quinol intermediate undergoes oxidation by an α-KG dependent NHI enzyme TropC to form a radical intermediate which undergoes a net ring rearrangement to form stipitatic aldehyde **1.59**.

Biocatalytic methods are poised to significantly expanded the repertoire of transformations possible in an organic chemist's toolbox, allowing greater access to chemical space than previously possible. This creates an incentive for academic and industrial labs to embrace biocatalytic methods. As interest in this field continues to grow, it will most certainly inform the retrosynthetic logic of modern organic synthesis and shape the next generation of methods.

1.5 Outlook and Conclusion

New technology and approaches in biocatalysis continue to pave the way for innovation and paint a bright future this field. Enzymatic catalysis has demonstrated utility in the construction of simple molecules and holds promise for expanding synthetic access

to new corners of chemical space. The rapid technological advances surrounding biocatalyst discovery, characterization and application naturally raises the question as to what comes next in the field. We anticipate that the amenability of biocatalysis to high-throughput experimentation will shape the application of enzymatic catalysis in synthesis. For example, we envision generation of compound libraries in plates will be possible through biocatalysis. Considering the benign nature of biocatalytic reactions, we anticipate biocatalytically generated compound libraries can be directly coupled with biological assays as well, matching the pace of compound generation with established high-throughput biological assays to ultimately accelerate drug discovery.[126, 127]

Continued progress in biocatalysis would benefit combinatorial platforms for the synthesis of small-molecule-based compound libraries. The idea of combinatorial biocatalysis platforms for library synthesis has been around since the early 2000s; however, its widespread adoption has been hindered by the lack of resources to identify and develop promiscuous catalytic enzymes.[128, 129] Combinatorial biocatalytic syntheses are now taking shape with recent advances in contemporary organic chemistry, synthetic biology, and bioinformatics. In addition, studies of enzyme cocktails have shown that biocatalysts can operate synergistically to complement each other's substrate scopes, creating useful catalyst mixtures to perform sequential chemical transformations.[130, 131] With this precedent, as well as equipment for high-throughput experimentation becoming more advanced and commonplace,[126] it seems only a matter of time before the high-throughput synthesis of vast and diverse small molecule libraries mediated by combinatorial biocatalysis is realized.

Without question, biocatalysis has become a valued approach in modern organic synthesis[126] and is a methodology we will rely heavily on as the need to develop green alternatives in chemistry grows.[17, 132] With the rapid advances in the field over the past few decades and the wealth of sequence data now widely available, biocatalytic methods are more accessible than ever before. As the global community adapts these techniques to their individual needs, new ideas and strategies will take hold and continue to push biocatalysis into the forefront of synthetic chemistry.

1.6 Dissertation Outline

The following Chapters will focus on my research regarding the chemoenzymatic synthesis of azaphilone natural products and the development of the platforms utilized to achieve these syntheses.

Chapter 2, "Sequence Similarity Network Guided Characterization of Flavin-Dependent Monooxygenases", summarizes our studies on characterizing several known and unknown FDMOs identified through an SSN constructed for this protein family by Dr. Attabey Rodríguez Benítez (Figure 1.7A). To enable the stereodivergent synthesis of azaphilone natural products, it was necessary to identify FDMOs with complementary stereoselectivity to those examined previously by our group. Dr. Benítez first constructed this SSN using the sequences of FDMOs TropB, AzaH, and SorbC as input sequences. With assistance from Dr. Summer Baker Dockrey, seven previously uncharacterized FDMOs and one characterized one, AfoD, were selected for further investigation to examine trends between the location of these sequences in the SSN and their site- and stereoselectivity in a model dearomatization reaction. Our findings support a link between

which cluster a sequence is located in the SSN and its site-selectivity in the dearomatization reaction, indicating this tool can be useful in identifying biocatalysts with a desired site-selectivity in this reaction. However, these investigations have shown that the stereoselectivity afforded by these enzymes is more finely controlled than can be predicted by the SSN: evidence suggests that two active site residues play a major role in the facial selectivity of the dearomzation.

Chapter 3, "Chemoenzymatic, Stereodivergent Synthesis of Azaphilone Natural Product" discusses our efforts to synthesize azaphilone natural products with opposing C7 configuration using complementary biocatalysts (Figure 1.7B). Once Dr. Baker Dockrey, Dr. Rodríguez Benítez, and I determined FDMOs AzaH and AfoD would be suitable catalysts for this project, we constructed a focused substrate library to identify

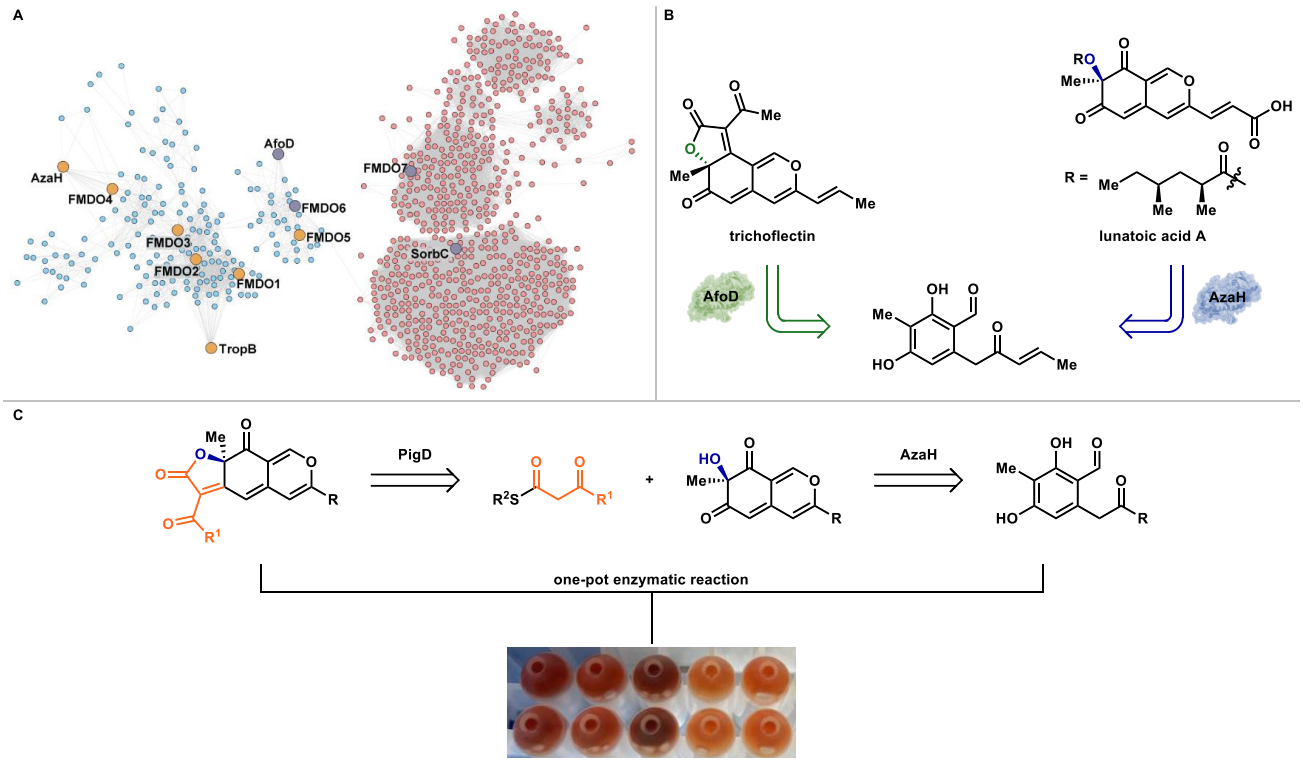


Figure 1.7. Visual overview of Chapters 2-4. A) Chapter 2, Sequence Similarity Network Guided Characterization of Flavin-Dependent Monooxygenases. B) Chapter 3, Stereodivergent, Chemoenzymatic Synthesis of Azaphilone Natural Products. C) Chapter 4, One-pot synthesis of tricyclic azaphilones using a dual-enzyme sequence

a suitable starting compound with functional handles for downstream chemistry. After we selected an orcinaldehyde bearing an enone group as the common intermediate, AzaH and AfoD were used to dearomatize this compound, ultimately providing access to both enantiomers of the natural product trichoflectin. Further characterization of these materials by Dr. Leo Joyce, Dr. Ren Wiscons, and myself provided evidence for the reassignment of the C7 configuration of the natural product from *S* to *R*, prompting me to investigate a related tricyclic azaphilone. Through comparisons to CD calculations provided by Dr. Joyce, my total synthesis of the natural product deflectin-1a allowed for structural revision of the deflectin series based on our findings, as well. Finally, I synthesized lunatoic acid A from the dearomatized intermediate accessed through AzaH, completing our stereodivergent syntheses.

I then pursued the synthesis of the other tricyclic azaphilone natural products and analogs thereof in **Chapter 4, "One-pot synthesis of azaphilone tricycles through sequential-addition biocatalysis"** with the help of Dr. Ye Wang. To begin the investigation of chemoenzymatic generation of azaphilone libraries in a well-plate format, I investigated two acyltransferase enzymes, PigD and CazE, for their ability to functionalize azaphilone bicycles. Experiments were first performed to examine the activity of these enzymes both in a stand-alone reaction and in a one-pot sequential set up with the FDMO, AzaH, performing an initial dearomatization reaction (Figure 1.7C). Though CazE did not demonstrate any activity, UPLC and LC-MS analysis of the crude extracts from reactions containing PigD confirmed the formation of the desired acylated material. Following reaction condition optimization and purification of PigD, a focused substrate library was screened against the developed one-pot conditions, revealing that

PigD exhibits promiscuity toward azaphilone substrates. In addition, initial spectroscopic data gathered by Dr. Wang and myself suggests that the acylated product intermediates in this screen undergo an unexpected cyclization, forming linear tricycles as opposed to the expected angular tricycles. We performed preparative scale reactions to provided access to the natural product, rubropunctatin, and two other linear tricycle analogs for NMR studies.

The last section, **Chapter 5, "Conclusions and Future directions"**, will provide a summary of the thesis work and a discussion of next steps and future experiments for the one-pot synthesis project.

1.7 References

(1) Bornscheuer, U. T.; Buchholz, K. Highlights in Biocatalysis – Historical Landmarks and Current Trends. *Eng.Life Sci.* **2005**, *5*, 309-323.
(2) Dominy, N. J. Ferment in the family tree. *Proc. Nat. Acad. Sci.* **2015,** *112*, 308-309.
(3) Chambers, P. J.; Pretorius, I. S. Fermenting knowledge: the history of winemaking, science and yeast research. *EMBO rep.* **2010**, *11*, 914-920.
(4) Copeland, R.A. A Brief History of Enzymology. Enzymes: A practical introduction to structure, mechanism, and data analysis, **2000**, 1-10.
(5) Hughes, G.; Lewis, J. C. Introduction: Biocatalysis in Industry. *Chem. Rev.* **2018**, *118*, 1-3.
(6) Devine, P. N.; Howard, R. M.; Kumar, R.; Thompson, M. P.; Truppo, M. D.; Turner, N. J. Extending the application of biocatalysis to meet the challenges of drug development. *Nat. Rev. Chem.* **2018**, *2*, 409-421.
(7) Truppo, M. D. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med. Chem. Lett.* **2017**, *8*, 476-480.
(8) Abdelraheem, E. M. M.; Busch, H.; Hanefeld, U.; Tonin, F. Biocatalysis explained: from pharmaceutical to bulk chemical production. *React. Chem. Eng.* **2019**, *4*, 1878-1894.
(9) Shin, J. S.; Kim, B. G.; Liese, A.; Wandrey, C. Kinetic resolution of chiral amines with omega-transaminase using an enzyme-membrane reactor. *Biotechnol. Bioeng.* **2001**, *73*, 179-87.
(10) Sharfuddin, M.; Narumi, A.; Iwai, Y.; Miyazawa, K.; Yamada, S.; Kakuchi, T.; Kaga, H. Lipase-catalyzed dynamic kinetic resolution of hemiaminals. *Tetrahedron Asymmetry* **2003**, *14*, 1581-1585.

(11) de Miranda, A. S.; Miranda, L. S. M.; de Souza, R. O. M. A. Lipases: Valuable catalysts for dynamic kinetic resolutions. *Biotechnol. Adv.* **2015**, *33*, 372-393.
(12) Koszelewski, D.; Lavandera, I.; Clay, D.; Guebitz, G. M.; Rozzell, D.; Kroutil, W. Formal Asymmetric Biocatalytic Reductive Amination. *Angew. Chem. Int. Ed.* **2008**, *47*, 9337-9340.
(13) Akoh, C. C.; Chang, S.-W.; Lee, G.-C.; Shaw, J.-F. Biocatalysis for the Production of Industrial Products and Functional Foods from Rice and Other Agricultural Produce. *J. Agric. Food Chem.* **2008**, *56*, 10445-10451.
(14) Huffman, M. A.; Fryszkowska, A.; Alvizo, O.; Borra-Garske, M.; Campos, K. R.; Canada, K. A.; Devine, P. N.; Duan, D.; Forstater, J. H.; Grosser, S. T. et al. Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science* **2019**, *366*, 1255-1259.
(15) Expanding biocatalysis for a sustainable future. *Nat. Catal.* **2020**, *3*, 179-180.
(16) Hayler, J. D.; Leahy, D. K.; Simmons, E. M. A Pharmaceutical Industry Perspective on Sustainable Metal Catalysis. *Organometallics* **2019**, *38*, 36-46.
(17) Chakrabarty, S.; Wang, Y.; Perkins, J. C.; Narayan, A. R. H., Scalable biocatalytic C–H oxyfunctionalization reactions. *Chem. Soc. Rev.* **2020**, *49*, 8137-8155.
(18) Chakrabarty, S.; Romero, E. O.; Pyser, J. B.; Yazarians, J. A.; Narayan, A. R. H. Chemoenzymatic Total Synthesis of Natural Products. *Acc. Chem. Res.* **2021**, *54*, 1374-1384.
(19) Lachance, H.; Wetzel, S.; Kumar, K.; Waldmann, H. Charting, Navigating, and Populating Natural Product Chemical Space for Drug Discovery. *J. Med. Chem.* **2012**, *55*, 5989-6001.
(20) Reetz, M. T. Biocatalysis in Organic Chemistry and Biotechnology: Past, Present, and Future. *J. Am. Chem. Soc.* **2013**, *135*, 12480-12496.
(21) Heckmann, C. M.; Paradisi, F. Looking Back: A Short History of the Discovery of Enzymes and How They Became Powerful Chemical Tools. *Chem. Cat. Chem.* **2020**, *12*, 6082-6102.
(22) Armstrong, E. F. Enzymes: A Discovery and its Consequences. *Nature* **1933**, *131*, 535-537.
(23) Fisher, E. The influence of configuration on enzyme activity (Translated from German). *Der. Dtsch. Chem. Ges.* **1894**, *27*, 2984-2993.
(24) Kohler, R. The Background to Eduard Buchner's Discovery of Cell-Free Fermentation. *J. Hist. Bio.* **1971**, *4*, 35-61.
(25) Sumner, J. B. The Isolation and Crystallization of the Enzyme Urease: Preliminary Paper. *J. Biol. Chem.* **1926**, *69*, 435-441.
(26) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737-738.
(27) Feiten, M. C.; Di Luccio, M.; Santos, K. F.; de Oliveira, D.; Oliveira, J. V. X-Ray Crystallography as a Tool to Determine Three-Dimensional Structures of Commercial Enzymes Subjected to Treatment in Pressurized Fluids. *Appl. Biochem. Biotechnol.* **2017**, *182*, 429-451.
(28) Kazlauskas, R. J.; Bornscheuer, U. T. Biotransformations with Lipases. *Biotechnology*, **1998**, 6-191.
(29) de Romo, A. C. Tallow and the Time Capsule: Claude Bernard's Discovery of the Pancreatic Digestion of Fat. *Hist. Philos. Life Sci.* **1989**, *11*, 253-274.

(30) Sym, E. A. Action of esterase in the presence of organic solvents. *Biochem. J.* **1936**, *30*, 609-617.
(31) Sperry, W. M.; Brand, F. C. A Study of Cholesterol Esterase in Liver and Brain. *J. Biol. Chem.* **1941**, *137*, 377-387.
(32) Liu, W.; Wang, P. Cofactor regeneration for sustainable enzymatic biosynthesis. *Biotechnol. Adv.* **2007**, *25*, 369-384.
(33) Richter, M. Functional diversity of organic molecule enzyme cofactors. *Nat. Prod. Rep.* **2013**, *30*, 1324-1345.
(34) Wong, C.-H.; Whitesides, G. M. Enzyme-catalyzed organic synthesis: NAD(P)H cofactor regeneration by using glucose-6-phosphate and the glucose-5-phosphate dehydrogenase from Leuconostoc mesenteroides. *J. Am. Chem. Soc.* **1981**, *103*, 4890-4899.
(35) Baker Dockrey, S. A.; Lukowski, A. L.; Becker, M. R.; Narayan, A. R. H. Biocatalytic site- and enantioselective oxidative dearomatization of phenols. *Nat. Chem.* **2018**, *10*, 119-125.
(36) Pyser, J. B.; Baker Dockrey, S. A.; Benítez, A. R.; Joyce, L. A.; Wiscons, R. A.; Smith, J. L.; Narayan, A. R. H. Stereodivergent, Chemoenzymatic Synthesis of Azaphilone Natural Products. *J. Am. Chem. Soc.* **2019**, *141*, 18551-18559.
(37) De Wildeman, S. M. A.; Sonke, T.; Schoemaker, H. E.; May, O. Biocatalytic Reductions: From Lab Curiosity to “First Choice”. *Acc. Chem. Res.* **2007**, *40*, 1260-1266.
(38) Wichmann, R.; Vasic-Racki, D. Cofactor Regeneration at the Lab Scale. *Adv. Biochem. Eng. Biotechnol.* **2005**, *92*, 225-260.
(39) Hughes, D. L. Biocatalysis in Drug Development—Highlights of the Recent Patent Literature. *Org. Process Res. Dev.* **2018**, *22*, 1063-1080.
(40) Cassimjee, K. E.; Branneby, C.; Abedi, V.; Wells, A.; Berglund, P. Transaminations with isopropyl amine: equilibrium displacement with yeast alcohol dehydrogenase coupled to in situ cofactor regeneration. *Chem. Comm.* **2010**, *46*, 5569-5571.
(41) Truppo, M. D.; Rozzell, J. D.; Moore, J. C.; Turner, N. J. Rapid screening and scale-up of transaminase catalysed reactions. *Org. Biomol. Chem.* **2009**, *7*, 395-398.
(42) Zachos, I.; Nowak, C.; Sieber, V. Biomimetic cofactors and methods for their recycling. *Curr. Opin. Chem. Biol.* **2019**, *49*, 59-66.
(43) Kelly, S. A.; Mix, S.; Moody, T. S.; Gilmore, B. F. Transaminases for industrial biocatalysis: novel enzyme discovery. *Appl. Microbiol. Biotechnol.* **2020,** *104* 4781-4794.
(44) Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J. et al. Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science* **2010**, *329*, 305-309.
(45) Slabu, I.; Galman, J. L.; Lloyd, R. C.; Turner, N. J. Discovery, Engineering, and Synthetic Application of Transaminase Biocatalysts. *ACS Catal.* **2017**, *7*, 8263-8284.

(46) Kelly, S. A.; Pohle, S.; Wharry, S.; Mix, S.; Allen, C. C. R.; Moody, T. S.; Gilmore, B. F. Application of ω-Transaminases in the Pharmaceutical Industry. *Chem. Rev.* **2018**, *118*, 349-367.
(47) Wandrey, C.; Liese, A.; Kihumbu, D. Industrial Biocatalysis: Past, Present, and Future. *Org. Process Res. Dev.* **2000**, *4*, 286-290.
(48) Chênevert, R.; Gagnon, R.; Desjardins, M.; Dickman, M.; Bureau, P.; Fortier, G.; Martin, R.; Létourneau, M.; Thiboutot, S.; Bel-Rhlid, R. Chemoenzymatic Synthesis of Natural Products and Bioactive Compounds. *Microbial Reagents in Organic Synthesis* **1992**, 135-147.
(49) McKean, I. J. W.; Hoskisson, P. A.; Burley, G. A. Biocatalytic Alkylation Cascades: Recent Advances and Future Opportunities for Late-Stage Functionalization. *ChemBioChem* **2020**, *21*, 2890-2897.
(50) Roumpeka, D. D.; Wallace, R. J.; Escalettes, F.; Fotheringham, I.; Watson, M. A Review of Bioinformatics Tools for Bio-Prospecting from Metagenomic Sequence Data. *Front. Genet.* **2017**, *8*, 1-10.
(51) Shendure, J.; Balasubramanian, S.; Church, G. M.; Gilbert, W.; Rogers, J.; Schloss, J. A.; Waterston, R. H. DNA sequencing at 40: past, present and future. *Nature* **2017**, *550*, 345-353.
(52) Packer, M. S.; Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **2015**, *16*, 379-394.
(53) The UniProt Consortium, UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **2017**, *45*, D158-D169.
(54) Sayers, E. W.; Cavanaugh, M.; Clark, K.; Ostell, J.; Pruitt, K. D.; Karsch-Mizrachi, I. GenBank. *Nucleic Acids Res.* **2019**, *47*, D94-D99.
(55) Gerlt, J. A, Tools and strategies for discovering novel enzymes and metabolic pathways. *Perspectives in Science* **2016**, *9*, 24-32.
(56) Naveed, M.; Chaudhry, Z.; Ali, Z.; Amjad, M.; Zulfiqar, F.; Numan, A. W. Annotation and curation of hypothetical proteins: prioritizing targets for experimental study. *Adv. Life Sci.* **2018**, *5*, 73-87.
(57) Sandoval, B. A.; Hyster, T. K. Emerging strategies for expanding the toolbox of enzymes in biocatalysis. *Curr. Opin. Chem. Biol.* **2020**, *55*, 45-51.
(58) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403-410.
(59) Boratyn, G. M.; Camacho, C.; Cooper, P. S.; Coulouris, G.; Fong, A.; Ma, N.; Madden, T. L.; Matten, W. T.; McGinnis, S. D.; Merezhuk, Y.; Raytselis, Y.; Sayers, E. W.; Tao, T.; Ye, J.; Zaretskaya, I. BLAST: a more efficient report with usability improvements. *Nucleic Acids Research* **2013**, *41*, W29–W33.
(60) Rodríguez Benítez, A.; Tweedy, S. E.; Baker Dockrey, S. A.; Lukowski, A. L.; Wymore, T.; Khare, D.; Brooks, C. L.; Palfey, B. A.; Smith, J. L.; Narayan, A. R. H. Structural Basis for Selectivity in Flavin-Dependent Monooxygenase-Catalyzed Oxidative Dearomatization. *ACS Catal.* **2019**, *9*, 3633-3640.
(61) Pearson, W. R., An introduction to sequence similarity ("homology") searching. *Curr. Protoc. Bioinf.* **2013**, *42*, 3.1.1-3.1.8.
(62) Madden, T. The BLAST Sequence Analysis Tool, *The NCBI Handbook*, **2013**, *2*, 361-370.

(63) Cai, X.-H.; Jaroszewski, L.; Wooley, J.; Godzik, A. Internal organization of large protein families: relationship between the sequence, structure, and function-based clustering. *Proteins* **2011**, *79*, 2389-2402.
(64) Rokas, A. Phylogenetic Analysis of Protein Sequence Data Using the Randomized Axelerated Maximum Likelihood (RAXML) Program. *Curr. Protoc. Mol. Biol.* **2011**, *96*, 19.11.1-19.11.14.
(65) Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* **2011**, *28*, 2731-2739.
(66) Hall, B. G. Building Phylogenetic Trees from Molecular Data with MEGA. *Mol. Biol. Evol.* **2013**, *30*, 1229-1235.
(67) Yates, A. D.; Achuthan, P.; Akanni, W.; Allen, J.; Alvarez-Jarreta, J.; Amode, M. R.; Armean, I. M.; Azov, A. G.; Bennett, R.; Bhai, J.; et al. *Nucleic Acids Res.* **2020**, *48*, D682-D688.
(68) Jones, C. M.; Stres, B.; Rosenquist, M.; Hallin, S. Phylogenetic Analysis of Nitrite, Nitric Oxide, and Nitrous Oxide Respiratory Enzymes Reveal a Complex Evolutionary History for Denitrification. *Mol. Biol. Evol.* **2008**, *25*, 1955-1966.
(69) Cavalcanti, J. H. F.; Esteves-Ferreira, A. A.; Quinhones, C. G. S.; Pereira-Lima, I. A.; Nunes-Nesi, A.; Fernie, A. R.; Araújo, W. L. Evolution and Functional Implications of the Tricarboxylic Acid Cycle as Revealed by Phylogenetic Analysis. *Genome Biol. Evol.* **2014**, *6*, 2830-2848.
(70) Siddiq, M. A.; Hochberg, G. K. A.; Thornton, J. W. Evolution of protein specificity: insights from ancestral protein reconstruction. *Curr. Opin. Struct. Biol.* **2017**, *47*, 113-122.
(71) Thornton, J. W. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **2004**, *5*, 366-375.
(72) Furukawa, R.; Toma, W.; Yamazaki, K.; Akanuma, S. Ancestral sequence reconstruction produces thermally stable enzymes with mesophilic enzyme-like catalytic properties. *Sci. Rep.* **2020**, *10*, 15493.
(73) O'Brien, P. J.; Herschlag, D. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **1999**, *6*, R91-R105.
(74) Rodríguez Benítez, A.; Narayan, A. R. H. Frontiers in Biocatalysis: Profiling Function across Sequence Space. *ACS Cent. Sci.* **2019,** *5*, 1747-1749.
(75) Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2015,** *1854*, 1019-1037.
(76) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome res.* **2003**, *13*, 2498-504.
(77) Atkinson, H. J.; Morris, J. H.; Ferrin, T. E.; Babbitt, P. C. Using Sequence Similarity Networks for Visualization of Relationships Across Diverse Protein Superfamilies. *PLOS ONE* **2009**, *4*, e4345.

(78) Fisher, B. F.; Snodgrass, H. M.; Jones, K. A.; Andorfer, M. C.; Lewis, J. C. Site-Selective C-H Halogenation Using Flavin-Dependent Halogenases Identified via Family-Wide Activity Profiling. *ACS Cent. Sci.* **2019**, *5*, 1844-1856.
(79) Wages, J. M. POLYMERASE CHAIN REACTION. *Encyclopedia of Analytical Science (Second Edition)*, Worsfold, P.; Townshend, A.; Poole, C., Eds. Elsevier: Oxford, **2005**, 243-250.
(80) Smalla, K.; Jechalke, S.; Top, E. M., Plasmid Detection, Characterization, and Ecology. *Microbiol. Spectr.* **2015**, *3*.
(81) Bruce Alberts, A. J., Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of the Cell, 4th edition*. 2002.
(82) Hughes, R. A.; Ellington, A. D. Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. *Cold Spring Harb. Perspect. Biol.* **2017**, *9*.
(83) Green, R.; Rogers, E. J. Transformation of chemically competent E. coli. *Methods Enzymol.* **2013**, *529*, 329-336.
(84) Wingfield, P. T. Overview of the purification of recombinant proteins. *Curr. Protoc. Protein Sci.* **2015**, *80*, 6.1.1-6.1.35.
(85) Pal, S. K.; Bandyopadhyay, S.; Ray, S. S. Evolutionary computation in bioinformatics: a review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **2006**, *36*, 601-615.
(86) Pengyi, Y.; Yee Hwa, Y.; Bing, B. Z.; Albert, Y. Z. A Review of Ensemble Methods in Bioinformatics. *Curr. Bioinf.* **2010**, *5*, 296-308.
(87) Wilding, M.; Scott, C.; Warden, A. C. Computer-Guided Surface Engineering for Enzyme Improvement. *Sci. Rep.* **2018**, *8*, 11998.
(88) Lowell, A. N.; DeMars, M. D.; Slocum, S. T.; Yu, F.; Anand, K.; Chemler, J. A.; Korakavi, N.; Priessnitz, J. K.; Park, S. R.; Koch, A. A. et al. Chemoenzymatic Total Synthesis and Structural Diversification of Tylactone-Based Macrolide Antibiotics through Late-Stage Polyketide Assembly, Tailoring, and C—H Functionalization. *J. Am. Chem. Soc.* **2017**, *139*, 7913-7920.
(89) Wang, S.-K.; Dai, C.-F.; Duh, C.-Y. Cytotoxic Pregnane Steroids from the Formosan Soft Coral Stereonephthya crystalliana. *J. Nat. Prod.* **2006**, *69*, 103-106.
(90) Wang, J.; Zhang, Y.; Liu, H.; Shang, Y.; Zhou, L.; Wei, P.; Yin, W.-B.; Deng, Z.; Qu, X.; Zhou, Q. A biocatalytic hydroxylation-enabled unified approach to C19-hydroxylated steroids. *Nat. Commun.* **2019**, *10*, 3378.
(91) Renata, H.; Zhou, Q.; Baran, P. S. Strategic Redox Relay Enables A Scalable Synthesis of Ouabagenin, A Bioactive Cardenolide. *Science* **2013**, *339*, 59-63.
(92) Renata, H.; Zhou, Q.; Dünstl, G.; Felding, J.; Merchant, R. R.; Yeh, C.-H.; Baran, P. S. Development of a Concise Synthesis of Ouabagenin and Hydroxylated Corticosteroid Analogues. *J. Am. Chem. Soc.* **2015**, *137*, 1330-1340.
(93) Wang, Y.; Ju, W.; Tian, H.; Tian, W.; Gui, J. Scalable Synthesis of Cyclocitrinol. *J. Am. Chem. Soc.* **2018**, *140*, 9413-9416.
(94) Lukowski, A. L.; Denomme, N.; Hinze, M. E.; Hall, S.; Isom, L. L.; Narayan, A. R. H. Biocatalytic Detoxification of Paralytic Shellfish Toxins. *ACS Chem. Biol.* **2019**, *14*, 941-948.
(95) Lukowski, A. L.; Ellinwood, D. C.; Hinze, M. E.; DeLuca, R. J.; Du Bois, J.; Hall, S.; Narayan, A. R. H. C–H Hydroxylation in Paralytic Shellfish Toxin Biosynthesis. *J. Am. Chem. Soc.* **2018**, *140*, 11863-11869.

(96) Lukowski, A. L.; Liu, J.; Bridwell-Rabb, J.; Narayan, A. R. H. Structural basis for divergent C–H hydroxylation selectivity in two Rieske oxygenases. *Nat. Commun.* **2020**, *11*, 2991.
(97) Lukowski, A. L.; Mallik, L.; Hinze, M. E.; Carlson, B. M.; Ellinwood, D. C.; Pyser, J. B.; Koutmos, M.; Narayan, A. R. H. Substrate Promiscuity of a Paralytic Shellfish Toxin Amidinotransferase. *ACS Chem. Biol.* **2020**, *15*, 626-631.
(98) Zhang, X.; King-Smith, E.; Dong, L.-B.; Yang, L.-C.; Rudolf, J. D.; Shen, B.; Renata, H. Divergent synthesis of complex diterpenes through a hybrid oxidative approach. *Science* **2020**, *369*, 799–806.
(99) Baud, D.; Saaidi, P.-L.; Monfleur, A.; Harari, M.; Cuccaro, J.; Fossey, A.; Besnard, M.; Debard, A.; Mariage, A.; Pellouin, V. et al. Synthesis of Mono- and Dihydroxylated Amino Acids with New α-Ketoglutarate-Dependent Dioxygenases: Biocatalytic Oxidation of C–H Bonds. *ChemCatChem* **2014**, *6*, 3012–3017.
(100) Baud, D.; Peruch, O.; Saaidi, P.-L.; Fossey, A.; Mariage, A.; Petit, J.-L.; Salanoubat, M.; Vergne-Vaxelaire, C.; de Berardinis, V.; Zaparucha, A. Biocatalytic Approaches towards the Synthesis of Chiral Amino Alcohols from Lysine: Cascade Reactions Combining alpha-Keto Acid Oxygenase Hydroxylation with Pyridoxal Phosphate-Dependent Decarboxylation. *Adv. Synth. Catal.* **2017**, *359*, 1563–1569.
(101) Zhang, X.; King-Smith, E.; Renata, H. Total Synthesis of Tambromycin by Combining Chemocatalytic and Biocatalytic C−H Functionalization. *Angew. Chem. Int. Ed.* **2018**, *57*, 5037–5041.
(102) Zetzsche, L. E.; Narayan, A. R. H. Broadening the scope of biocatalytic C–C bond formation. *Nat. Rev. Chem.* **2020**, *4*, 334–346.
(103) Nakamura, H.; Schultz, E. E.; Balskus, E. P. A new strategy for aromatic ring alkylation in cylindrocyclophane biosynthesis. *Nat. Chem. Biol.* **2017**, *13*, 916–921.
(104) Schultz, E. E.; Braffman, N. R.; Luescher, M. U.; Hager, H. H.; Balskus, E. P. Biocatalytic Friedel–Crafts Alkylation Using a Promiscuous Biosynthetic Enzyme. *Angew. Chem. Int. Ed.* **2019**, *58*, 3151–3155.
(105) Lau, W.; Sattely, E. S. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science* **2015**, *349*, 1224–1228.
(106) Chang, W.-c.; Yang, Z.-J.; Tu, Y.-H.; Chien, T.-C., Reaction Mechanism of a Nonheme Iron Enzyme Catalyzed Oxidative Cyclization via C–C Bond Formation. *Org. Lett.* **2019**, *21*, 228–232.
(107) Lazzarotto, M.; Hammerer, L.; Hetmann, M.; Borg, A.; Schmermund, L.; Steiner, L.; Hartmann, P.; Belaj, F.; Kroutil, W.; Gruber, K. et al. Chemoenzymatic Total Synthesis of Deoxy-, epi-, and Podophyllotoxin and a Biocatalytic Kinetic Resolution of Dibenzylbutyrolactones. *Angew. Chem. Int. Ed.* **2019**, *58*, 8226–8230.
(108) Li, J.; Zhang, X.; Renata, H. Asymmetric Chemoenzymatic Synthesis of (−)-Podophyllotoxin and Related Aryltetralin Lignans. *Angew. Chem. Int. Ed.* **2019**, *58*, 11657–11660.
(109) Mazzaferro, L. S.; Hüttel, W.; Fries, A.; Müller, M. Cytochrome P450-Catalyzed Regio- and Stereoselective Phenol Coupling of Fungal Natural Products. *J. Am. Chem. Soc.* **2015**, *137*, 12289–12295.

(110) Präg, A.; Grüning, B. A.; Häckh, M.; Lüdeke, S.; Wilde, M.; Luzhetskyy, A.; Richter, M.; Luzhetska, M.; Günther, S.; Müller, M. Regio- and Stereoselective Intermolecular Oxidative Phenol Coupling in Streptomyces. *J. Am. Chem. Soc.* **2014**, *136*, 6195–6198.
(111) Obermaier, S.; Müller, M. Biaryl-Forming Enzymes from Aspergilli Exhibit Substrate-Dependent Stereoselectivity. *Biochemistry* **2019**, *58*, 2589–2593.
(112) Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science* **2013**, *339*, 307–310.
(113) Coelho, P. S.; Wang, Z. J.; Ener, M. E.; Baril, S. A.; Kannan, A.; Arnold, F. H.; Brustad, E. M. A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. *Nat. Chem. Biol.* **2013**, *9*, 485–487.
(114) Bordeaux, M.; Tyagi, V.; Fasan, R. Highly Diastereoselective and Enantioselective Olefin Cyclopropanation Using Engineered Myoglobin-Based Catalysts. *Angew. Chem. Int. Ed.* **2015**, *54*, 1744–1748.
(115) Bajaj, P.; Sreenilayam, G.; Tyagi, V.; Fasan, R. Gram-Scale Synthesis of Chiral Cyclopropane-Containing Drugs and Drug Precursors with Engineered Myoglobin Catalysts Featuring Complementary Stereoselectivity. *Angew. Chem. Int. Ed.* **2016**, *55*, 16110–16114.
(116) Chandgude, A. L.; Ren, X.; Fasan, R. Stereodivergent Intramolecular Cyclopropanation Enabled by Engineered Carbene Transferases. *J. Am. Chem. Soc.* **2019**, *141*, 23, 9145–9150110.
(117) Chen, K.; Huang, X.; Kan, S. B. J.; Zhang, R. K.; Arnold, F. H. Enzymatic construction of highly strained carbocycles. *Science* **2018**, *360*, 71–75.
(118) Zwick, C. R.; Renata, H. Remote C–H Hydroxylation by an α-Ketoglutarate-Dependent Dioxygenase Enables Efficient Chemoenzymatic Synthesis of Manzacidin C and Proline Analogs. *J. Am. Chem. Soc.* **2018**, *140*, 1165–1169.
(119) Zwick, C. R.; Renata, H. Evolution of Biocatalytic and Chemocatalytic C–H Functionalization Strategy in the Synthesis of Manzacidin C. *J. Org. Chem.* **2018**, *83*, 7407–7415.
(120) Doyon, T. J.; Perkins, J. C.; Baker Dockrey, S. A.; Romero, E. O.; Skinner, K. C.; Zimmerman, P. M.; Narayan, A. R. H. Chemoenzymatic o-Quinone Methide Formation. *J. Am. Chem. Soc.* **2019**, *141*, 20269–20277.
(121) Takeda, R.; Abe, H.; Shibata, N.; Moriwaki, H.; Izawa, K.; Soloshonok, V. A. Asymmetric synthesis of α-deuterated α-amino acids. *Org. Biomol. Chem.* **2017**,*15*, 6978–6983.
(122) Chatterjee, B.; Krishnakumar, V.; Gunanathan, C. Selective α-Deuteration of Amines and Amino Acids Using D_2O. *Org. Lett.* **2016**, *18*, 5892–5895.
(123) McLaughlin, M.; Kong, J.; Belyk, K. M.; Chen, B.; Gibson, A. W.; Keen, S. P.; Lieberman, D. R.; Milczek, E. M.; Moore, J. C.; Murray, D.; Peng, F.; Qi, J.; Reamer, R. A.; Song, Z. J.; Tan, L.; Wang, L.; Williams, M. J. Enantioselective Synthesis of 4′-Ethynyl-2-fluoro-2′-deoxyadenosine (EFdA) via Enzymatic Desymmetrization. *Org. Lett.* **2017**, *19*, 926–929.
(124) McKinnie, S. M. K.; Miles, Z. D.; Jordan, P. A.; Awakawa, T.; Pepper, H. P.; Murray, L. A. M.; George, J. H.; Moore, B. S. Total Enzyme Syntheses of Napyradiomycins A1 and B1. *J. Am. Chem. Soc.* **2018**, *140*, 17840–17845.

(125) Tyler J., D.; Kevin, S.; Di, Y.; Leena, M.; Troy, W.; Markos, K.; Paul M., Z.; Alison, N. Radical Tropolone Biosynthesis. *ChemRxiv* Preprint, 2020, https://chemrxiv.org/articles/preprint/Radical_Tropolone_Biosynthesis/12780044/1.
(126) Mennen, S. M.; Alhambra, C.; Allen, C. L.; Barberis, M.; Berritt, S.; Brandt, T. A.; Campbell, A. D.; Castañón, J.; Cherney, A. H.; Christensen, M. et al. The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Org. Process Res. Dev.* **2019**, *23*, 1213–1242.
(127) Dandapani, S.; Rosse, G.; Southall, N.; Salvino, J. M.; Thomas, C. J. Selecting, Acquiring, and Using Small Molecule Libraries for High-Throughput Screening. *Curr. Protoc. Chem. Biol.* **2012**, *4*, 177–191.
(128) Rich, J. O.; Michels, P. C.; Khmelnitsky, Y. L. Combinatorial biocatalysis. *Curr. Opin. Chem. Biol.* **2002**, *6*, 161–167.
(129) Altreuter, D. H.; Clark, D. S. Combinatorial biocatalysis: taking the lead from Nature. *Currn. Opin. Biotechnol.* **1999**, *10*, 130–136.
(130) Agrawal, R.; Semwal, S.; Kumar, R.; Mathur, A.; Gupta, R. P.; Tuli, D. K.; Satlewal, A. Synergistic Enzyme Cocktail to Enhance Hydrolysis of Steam Exploded Wheat Straw at Pilot Scale. *Front. Energy Res.* **2018**, 6.
(131) Poppe, J. K.; Matte, C. R.; de Freitas, V. O.; Fernandez-Lafuente, R.; Rodrigues, R. C.; Záchia Ayub, M. A. Enzymatic synthesis of ethyl esters from waste oil using mixtures of lipases in a plug-flow packed-bed continuous reactor. *Biotechnol. Progress* **2018**, *34*, 952–959.
(132) Sheldon, R. A.; Woodley, J. M. Role of Biocatalysis in Sustainable Chemistry. *Chem. Rev.* **2018**, *118*, 801–838.

Chapter 2: Sequence Similarity Network Guided Characterization of Flavin-Dependent Monooxygenases

Summary

Selective access to a targeted stereoisomer is often critical in synthesis of biologically active molecules. Whereas small-molecule reagents and catalysts are routinely available as either enantiomer, mirror-image biocatalysts are not readily accessible, creating a challenge in the application of biocatalysis in synthesis. This difficulty is highlighted when considering the synthesis of members of the class of natural products known as azaphilones, which posses a hydroxyl group bearing either configuration at C7. To demonstrate the utility of biocatalysts to construct either enantiomer of the azaphilone core selectively, we sought to identify homologous enzymes capable of performing the desired oxidative dearomatization chemistry with complementary selectivity on the same aromatic precursor. However, despite the abundance of sequence information available in online databases, identifying reactivity trends between previously uncharacterized enzymes poses a significant challenge. Herein, we report the use of a bioinformatic tool

called a sequence similarity network (SSN), constructed by Dr. Attabey Rodríguez Benítez, to identify FDMOs providing complimentary selectivity in their activity with a model substrate. The constructed SSN was successful in predicting the site selectivity of chosen FDMOs in their dearomatization reaction based on their cluster location and provided evidence suggesting that the stereoselectivity of these enzymes is more finely controlled than can be predicted by their SSN location.

2.1 Introduction

Biocatalytic approaches can provide numerous advantages in synthesis. These include high levels of selectivity, mild reaction conditions, and opportunities for cascade reactions that enable the rapid construction of complex molecules.[1] These attractive qualities have drawn the attention of the synthetic community, sparking growth in the application of biocatalysis in the synthesis of bioactive compounds.[2-7] When a particular biological activity motivates the synthesis of a target molecule, the ability to access individual isomers of the final compound is critical.[8] Traditional chemical strategies for selective synthesis can provide tools to arrive at a predicted selectivity outcome. For example, the desired enantiomer can be accessed by employing chiral ligands that can be obtained from commercial sources or synthesized as either enantiomer.[9, 10] This strategy is not currently applicable to biocatalytic approaches, as methods for generation of mirror-image proteins are in the early stages of development.[11, 12] Protein engineering is commonly employed to address this problem (Figure 2.1A); however, there is no unified approach for enabling access to diverse selectivity outcomes biocatalytically. Here, we seek to address this drawback by employing an alternative approach to identifying panels

of catalysts which deliver complementary selectivity outcomes in a synthetically valuable transformation.

Previously, our group has demonstrated the utility of select biocatalysts for site- and stereoselective oxidative dearomatization of resorcinol compounds.[13] The flavin-dependent enzymes used in this study provide distinct advantages over small-molecule reagents and catalysts for asymmetric dearomatization, as these biocatalysts are perfectly site-selective and avoid overoxidation that leads to undesired byproducts.[13-15] However, the set of biocatalysts reported previously hydroxylate to afford products with only the *R*-configuration at C3 (see (*R*)-**2.12**, Figure 2.1B) limiting the potential

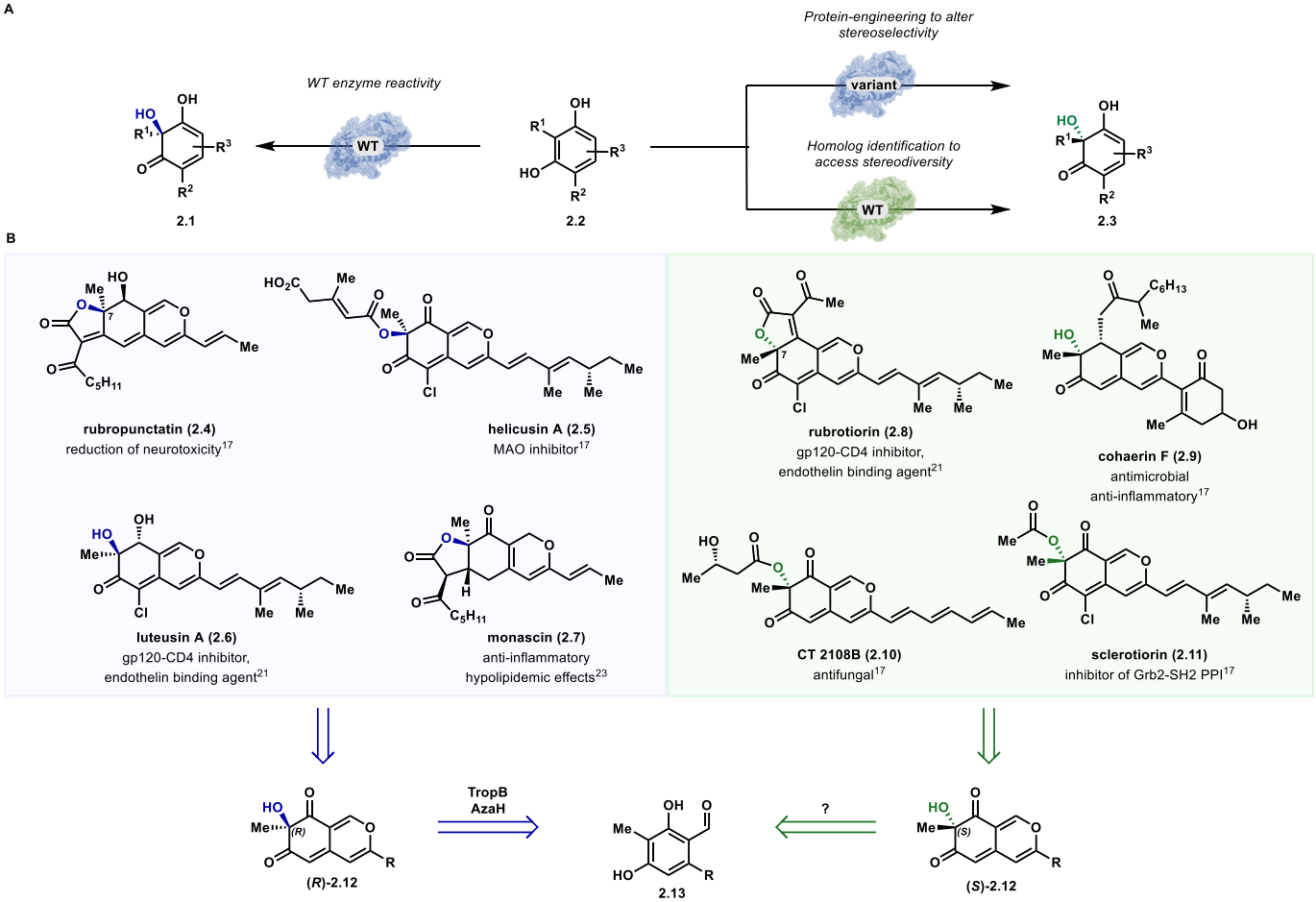


Figure 2.1. A) Approaches to enabling biocatalytic stereodiversity. B) Biologically active azaphilone natural products feature both the *R*- and *S*-configuration at the C7 stereocenter.[17,21,23] Biocatalytic oxidative dearomatization gives access to either enantiomer via AzaH or AfoD mediated oxidation.

applications of this method. To access the enantiomeric set of products, we considered two approaches: protein engineering to achieve a switch in stereoselectivity or use of a natural enzyme with complementary selectivity.

Over the last decade, there has been exponential growth in the number of annotated protein sequences available, due in part to the continuing decrease in the price of sequencing.[16] This sequencing revolution coupled with improved bioinformatic tools for predicting enzyme class and function provides an alternative to protein engineering campaigns, making it increasingly possible to identify a natural enzyme that can perform a given biocatalytic reaction with the desired selectivity (Figure 2.1A). Based on the number of natural products that could biosynthetically arise through enantioselective hydroxylative dearomatization, we anticipated it would be possible to identify enzymes capable of providing the complementary stereoisomer to the flavin-dependent monooxygenases TropB and AzaH employed in our previous study (Figure 2.1B).[13] Once stereocomplementary catalysts have been identified, we envision that this suite of biocatalysts will enable the stereodivergent, chemoenzymatic synthesis of high-value molecules.

To showcase this strategy, we targeted the azaphilone family of natural products. These compounds are characterized by an oxygenated pyranoquinone bicyclic core bearing a single tetrasubstituted carbon (Figure 2.1B).[17] Isolated from fungal sources, this large family of natural products is known to contain a diverse array of structural features that impart a wide range of biological properties including anticancer,[18-20] antiviral,[21] and anti-inflammatory activities.[20, 22] For example, luteusin A (**2.6**) and rubrorotiorin (**2.8**) were found to inhibit the binding of the HIV surface glycoprotein gp120 to the human CD4

protein, making these natural products potential starting points for the development of therapeutic agents against the virus.[21] In addition, monascin (**2.7**), a tricyclic azaphilone natural product, has been shown to down-regulate steatohepatitis in a mouse model, indicating that these secondary metabolites have potential as therapeutics for non-alcoholic fatty liver disease.[23] As illustrated in Figure 2.1B, azaphilone natural products can contain either the *R*- or *S*-configuration at the C7-position. Several cases of epimeric azaphilones are known, in which each C7-epimer is produced by a distinct fungal source.[24, 25] The pharmaceutical potential of these molecules has been demonstrated through initial *in vitro* and cell-based assays of isolated natural products;[26-28] however, to gain a more comprehensive understanding of the therapeutic potential of these molecules, the challenge of constructing the densely functionalized core and C7-configuration for exploration of this stereocenter's impact on biological activity must be addressed.[24, 25, 29]

2.2 Constructing an SSN of the FDMO PFam

To survey the selectivity of flavin-dependent monooxygenases related by sequence, Dr. Rodríguez Benítez first constructed a sequence similarity network (SSN) of flavin-dependent monooxygenases (Figure 2.2A).[30-32] SSNs are visual representations of the relatedness of protein sequences, which cluster based on similarity thresholds. This approach can enable the rapid and logical investigation of proteins likely to demonstrate similar reactivity and/or selectivity, dramatically reducing the time and search space required during screening, considerations critical to improving the compatibility of biocatalytic approaches and traditional synthetic strategies. The protein family

(Pfam01494) of flavin-dependent monooxygenases includes over 45,000 enzymes, which are included in the full SSN (Supporting Information Figure 2.6). This data set was truncated by limiting the search to edges possessing an alignment score greater than 110, which returned 1,211 sequences (Figure 2.2A). She noted that the enzymes we had previously investigated, AzaH, TropB, and SorbC, were each located in distinct groups or "clusters" within the network, and that a fourth tight clustering of sequences formed between the TropB and SorbC clusters containing an enzyme associated with asperfuranone biosynthesis, AfoD.[33-35] Previous in vivo studies[35] and bioinformatic analysis[34] indicated that AfoD is responsible for oxidative dearomatization of an asperfuranone precursor with the same site-selectivity as TropB and AzaH;[34] however, the absolute configuration of asperfuranone suggests that AfoD carries out this transformation with the opposite facial selectivity.[33] We hypothesized that sequences with high similarity to tropB, azaH, afoD, and sorbC would encode for enzymes that possess the ability to carry out oxidative dearomatization reactions, and further questioned if trends in site- and stereoselectivity could be predicted on the basis of sequence.

2.3 Characterizing the Selectivity of Unknown FDMOs

To test this hypothesis, seven sequences were selected on the basis of proximity to either TropB, AzaH, AfoD, or SorbC in the SSN by Dr. Baker Dockrey, Dr. Rodríguez Benítez, and myself (labeled FDMO1−7 in Figure 2.2A). Synthetic genes corresponding to the four known enzymes and FDMO1−7 were transformed into E. coli BL21(DE3) cells. Under standard expression conditions, Dr. Rodríguez Benítez obtained nine proteins successfully, whereas FDMO1 and FDMO3 proved insoluble under these conditions

(Figure 2.2B). The reactivity of the nine soluble enzymes was evaluated by Dr. Baker Dockrey and I with two substrates, S1 and S14. Gratifyingly, all nine enzymes showed activity with one of the two model substrates, eight of which displayed sufficiently high activity for product isolation from preparative-scale reactions (Figure 2.2B). A single product was obtained from each biocatalytic reaction, and a strong trend in site-selectivity was clear from these data: FDMO7, the enzyme most similar to SorbC, afforded C5-hydroxylated product (see **2.16**), whereas the remaining enzymes more similar to TropB, AzaH, and AfoD delivered C3-hydroxylated products (see **2.15**). However, the trend between location on the SSN (Figure 2.2A) and stereoselectivity was not apparent. Dr. Rodríguez Benítez demonstrated that increasing the alignment score from 110 to 150

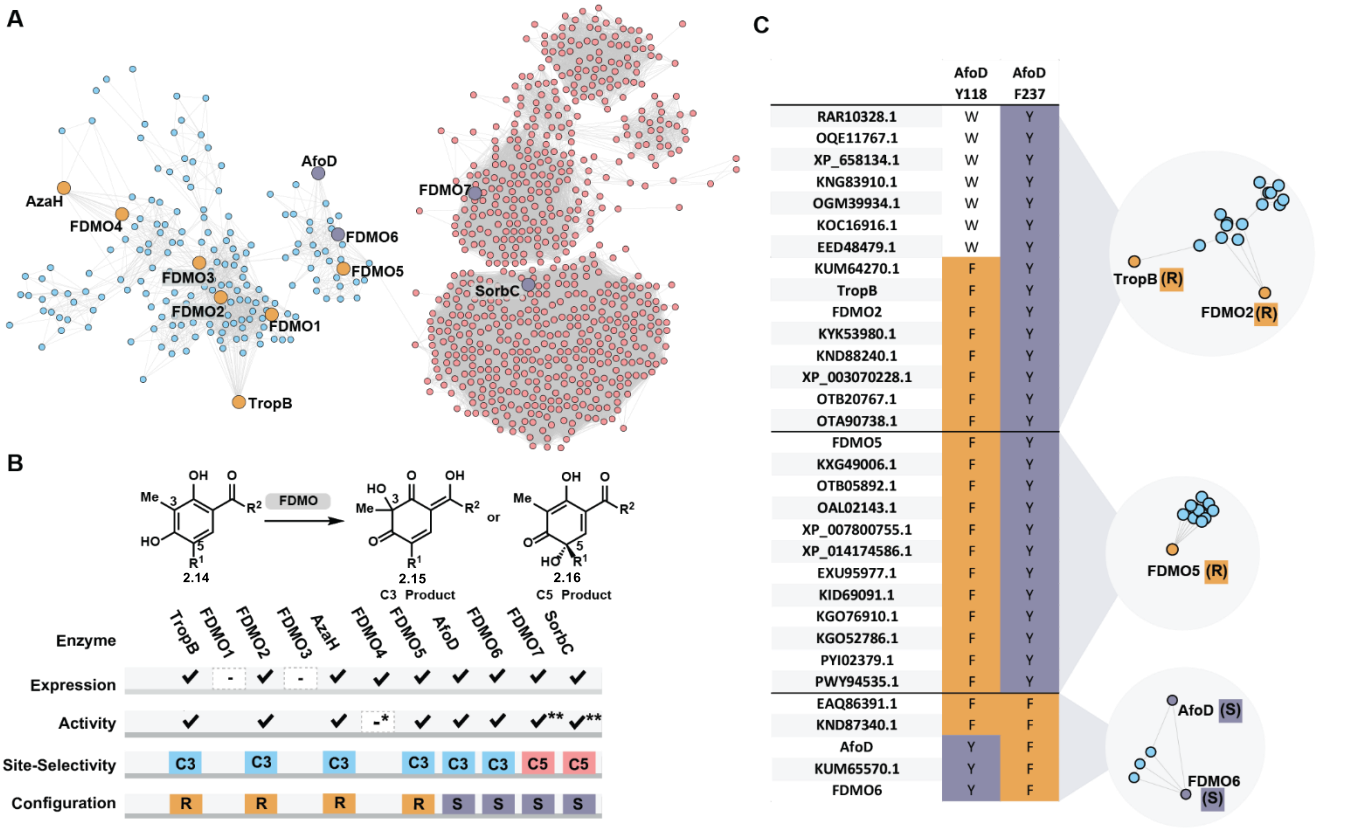


Figure 2.2. A) Sequence similarity network (SSN) of FAD-dependent monooxygenases (Pfam01494) using a sequence alignment score of 110, made by Dr. Attabey Rodríguez Benítez . B) Results of expression, activity with model substrate S1 or S3**, site- and stereoselectivity of enzymes chosen from the SSN in panel A. *FDMO4 demonstrated <10% conversion by UPLC with substrate S1. C) Selected clusters from a more stringent SSN generated with an alignment score of 150 and corresponding analysis of the multisequence alignment of each cluster.

produced an SSN provided greater insight into the relationships between sequences in this family (see Figure 2.7 for full SSN). Notably, previously clustered sequences associated with divergent selectivity, such as AfoD and FDMO5, now clearly separated into distinct clusters (see Figure 2.2C). Analysis of sequence alignments between each of these clusters revealed conserved residues at positions 118 and 237 (AfoD numbering). In general, catalysts that generate products with the R-configuration possess a tyrosine at position 237 and an aromatic residue such as phenyalanine or tryptophan at position 118. In contrast, these conserved residues switch positions in biocatalysts that afford products with the S-configuration. Dr. Rodríguez Benítez previously demonstrated that a two-point coordination of the phenolate substrate by Tyr237 and an arginine residue is critical for positioning the substrate within the active site in the R-selective enzyme, TropB.[36, 37] To probe the role of Tyr118 in the stereoselectivity provided by AfoD, she generated the AfoD Y118F variant. This was reacted with resorcinol **2.S1** to afford dearomatized product in 53:47 (S:R) er (Figure 2.12). These data provide additional evidence for the importance of this tyrosine residue in controlling stereoselectivity in these biocatalytic reactions and support the consideration of this sequence feature in choosing biocatalysts for a specific synthetic application. Obtaining additional structure and function data on these proteins will aid in demystifying the elements that contribute to the precise stereocontrol exhibited by these catalysts. Beyond the importance of Tyr position in the active site, it is clear that other mechanisms for stereocontrol have evolved in this class of enzymes. For example, Dr. Baker Dockrey showed previously that catalysts AzaH and SorbC break from this Tyr control mechanism, and, in the case of SorbC, she has proposed an alternative mechanism for control of substrate position in the active site.[38]

Ultimately, this survey of sequence space surrounding known enzymes provides a greater understanding of the sequence features that can predict site- and stereoselectivity and has increased the number of biocatalysts vetted for this transformation. With catalysts capable of delivering enantiomeric products in hand, we chose to pursue an enantiodivergent synthetic strategy with AzaH and AfoD based on the robust expression and reactivity of these enzymes, in addition to the excellent stereoselectivity of each catalyst. Under reaction conditions to dearomatize model substrate **2.S1** using AzaH and AfoD, enantiomeric bicycles (R)- and (S)- **2.S2** were accessed in >99% ee and 98% ee, respectively (Figure 2.12).

2.4 Conclusion

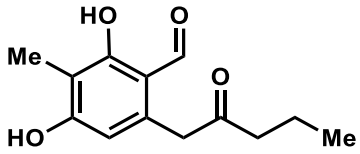
Given the abundance of biological activity reported for the azaphilone natural product class, we sought to develop a method for their selective synthesis using a biocatalytic approach mediated by FDMOs. To provide a means to predict the site- and stereoselectivity of FDMOs in a model dearomatization reaction, Dr. Rodríguez Benítez constructed an SSN of PFam01494. Nine sequences were ultimately chosen from this SSN for further characterization based on their cluster location. After successful expression, it was found that all nine catalysts demonstrated activity in a model oxidative dearomatization reaction, with eight of the enzymes producing sufficient product yields for further characterization. We identified a trend between cluster location and site-selectivity of these biocatalysts, suggesting that this bioinformatic tools can be used to predict this trait in previously uncharacterized protein sequences. However, no clear trend between SSN location and stereoselectivity was determined, demonstrating that this trait

is more finely controlled. Examining the PFam01494 SSN with an increased threshold scor0065, alongside follow-up experiments with an AFoD variant, suggests a link between the identity of two key active site residues and the stereochemical outcome of the dearomatization reaction. These findings help to better understand this class of enzymes and enable their use in biocatalytic syntheses.

2.5 Experimental

I. Substrate Synthesis

All reagents were used as received unless otherwise noted. Reactions were carried out under a nitrogen atmosphere using standard Schlenck techniques unless otherwise noted. Solvents were degassed and dried over aluminum columns on an MBraun solvent system (Innovative Technology, inc., Model PS-00-3). Reactions were monitored by thin layer chromatography using Millipore 60 F_{254} precoated silica TLC plates (0.25 mm) that were visualized using UV, *p*-anisaldehyde, CAM, DNP, or bromocresol stain. Flash column chromatography was performed using Machery-Nagel 60 μm (230-400 mesh) silica gel. All compounds purified by column chromatography were sufficiently pure for use in further experiments unless otherwise indicated. ^{1}H and ^{13}C NMR spectra were obtained in $CDCl_3$ at rt (25 °C), unless otherwise noted, on Varian 400 MHz or Varian 600 MHz spectrometers. Chemical shifts of ^{1}H NMR spectra were recorded in parts per million (ppm) on the δ scale. High resolution electrospray mass spectra were obtained on an Agilent HPLC-TOF at the University of Michigan Life Sciences Institute.



2,4-Dihydroxy-3-methyl-6-(2-oxopentyl)benzaldehyde (2.S1)
Prepared as previously reported by Baker Dockrey *et al*.[13]

II. Plasmids and Proteins

Plasmids: The plasmid encoding *azaH* (G3XMC2.1), in a modified pET28 vector to afford protein with both C- and N-terminal 6 x His-tags, was a generous gift from Professor Yi Tang at the University of California, Los Angeles.[19] The plasmid encoding *afoD* was synthesized by GeneArt and cloned into a pET21a vector. The plasmids encoding for *fdmo1-7* were synthesized by Twist Bioscience and cloned into a pET28a vector.

Non-optimized *azaH* Sequence

ATGAGTACAGACTCGATCGAAGTTGCCATTATAGGCGCCGGGATCACGGGAATCA
CCCTGGCCCTGGGCCTCCTGTCTCGCGGCATTCCCGTCCGCGTCTACGAGCGAG
CCCGCGACTTTCACGAAATTGGAGCCGGTATCGGTTTCACCCCCAACGCCGAATG
GGCGATGAAAGTCGTCGACCCGCGCATTCAAGCTGCTTTCAAACGCGTCGCTACC
CCCAATGCCTCCGACTGGTTCCAGTGGGTGGACGGATTCAACGAGTCCGGTACCG
ACCCGCGCGAGACCGAGGAACAGCTACTCTTCAAGATCTACCTCGGCGAGCGTG
GATTTGAGGGCTGCCACCGTGCCGACTTCCTAGGTGAGCTGGCACGTCTACTACC
GGAAGGTGTGGTGACATTCCAGAAGGCGCTGGATACCGTGGAGCCTGCAGCAGA
TAATAGCCTCGGCCAGCTTCTTCGATTCCAAGATGGCACGACAGCTACCGCCCAC
GCGGTGATCGGCTGCGATGGCATTCGGTCGCGCGTTCGTCAGATCCTCCTAGGTG
AAGACCATCCGACAGCATCAGCCCATTACAGTCATAAATATGCAGCACGCGGCCTT
ATTCCCATGGACCGCGCCCGGGAGGCGCTGGGCGAAGATAAAGTGGCGACACGC
TTCATGCATCTCGGTCCGGATGCCCATGCCCTGACCTTCCCCGTTAGCCATGGGT
CCTTGTTGAACGTCGTCGCCTTCGTCACGGACCCTAACCCTTGGCCATATGCTGAT
CGCTGGACGGCGCAGGGGCCCAAGAAAGACGTGACGGCTGCCTTTTCCCGCTTT
GGTCCGACCATGCGCACCATAATTGACCTCTTGCCTGATCCTATTGATCAATGGGC
CGTTTTTGATACATACGACCATCCCCCAAATACGTATTCCCGGGGAGCTGTCTGTA
TAGCAGGGGATGCTGCTCATGCCGCGGCTCCGCATCACGGTGCAGGTGCAGGTT
GTGGTGTGGAAGACGCGGCTGTGCTGTGCGCTGTGCTTCATATGGCTGCGAAAAA
AGTTAACACCGCAAAAACTGGTTCTGAGGGGAAAGCCGCTCTTATCACGGCCGCA
TTCGAAACCTATGATTCGGTTTGTCGCGAGCGTGCGCAGTGGCTGGTGGAAAGTA
GTCGCGTTATCGGTAATCTGTATGAGTGGCAGGATAAGGAGGTAGGGTCGGATGC
TTCCAGGTGCCACGATGAGGTGTATTGGCGCTCTCATCGCATTTGGGACTATGATA
TTGATGCGATGATGAGAGAGACAGCTGAGGTGTTTGAGGCGCAGGTAGCTGGGGT
GGCGAGAAAT

AzaH Protein Sequence

MSTDSIEVAIIGAGITGITLALGLLSRGIPVRVYERARDFHEIGAGIGFTPNAEWAMKVVD
PRIQAAFKRVATPNASDWFQWVDGFNESGTDPRETEEQLLFKIYLGERGFEGCHRAD
FLGELARLLPEGVVTFQKALDTVEPAADNSLGQLLRFQDGTTATAHAVIGCDGIRSRVR
QILLGEDHPTASAHYSHKYAARGLIPMDRAREALGEDKVATRFMHLGPDAHALTFPVS
HGSLLNVVAFVTDPNPWPYADRWTAQGPKKDVTAAFSRFGPTMRTIIDLLPDPIDQWA
VFDTYDHPPNTYSRGAVCIAGDAAHAAAPHHGAGAGCGVEDAAVLCAVLHMAAKKVN
TAKTGSEGKAALITAAFETYDSVCRERAQWLVESSRVIGNLCHDEVYWRSHRIWDYDI
DAMMRETAEVFEAQVAGVARN

Codon-Optimized *afoD* Sequence

ATGAGTACAGACTCGATCGAAGTTGCCATTATAGGCGCCGGGATCACGGGAATCA
CCCTGGCCCTGGGCCTCCTGTCTCGCGGCATTCCCGTCCGCGTCTACGAGCGAG
CCCGCGACTTTCACGAAATTGGAGCCGGTATCGGTTTCACCCCCAACGCCGAATG
GGCGATGAAAGTCGTCGACCCGCGCATTCAAGCTGCTTTCAAACGCGTCGCTACC
CCCAATGCCTCCGACTGGTTCCAGTGGGTGGACGGATTCAACGAGTCCGGTACCG
ACCCGCGCGAGACCGAGGAACAGCTACTCTTCAAGATCTACCTCGGCGAGCGTG
GATTTGAGGGCTGCCACCGTGCCGACTTCCTAGGTGAGCTGGCACGTCTACTACC
GGAAGGTGTGGTGACATTCCAGAAGGCGCTGGATACCGTGGAGCCTGCAGCAGA
TAATAGCCTCGGCCAGCTTCTTCGATTCCAAGATGGCACGACAGCTACCGCCCAC
GCGGTGATCGGCTGCGATGGCATTCGGTCGCGCGTTCGTCAGATCCTCCTAGGTG
AAGACCATCCGACAGCATCAGCCCATTACAGTCATAAATATGCAGCACGCGGCCTT
ATTCCCATGGACCGCGCCCGGGAGGCGCTGGGCGAAGATAAAGTGGCGACACGC
TTCATGCATCTCGGTCCGGATGCCCATGCCCTGACCTTCCCCGTTAGCCATGGGT
CCTTGTTGAACGTCGTCGCCTTCGTCACGGACCCTAACCCTTGGCCATATGCTGAT
CGCTGGACGGCGCAGGGGCCCAAGAAAGACGTGACGGCTGCCTTTTCCCGCTTT
GGTCCGACCATGCGCACCATAATTGACCTCTTGCCTGATCCTATTGATCAATGGGC
CGTTTTTGATACATACGACCATCCCCCAAATACGTATTCCCGGGGAGCTGTCTGTA
TAGCAGGGGATGCTGCTCATGCCGCGGCTCCGCATCACGGTGCAGGTGCAGGTT
GTGGTGTGGAAGACGCGGCTGTGCTGTGCGCTGTGCTTCATATGGCTGCGAAAAA
AGTTAACACCGCAAAAACTGGTTCTGAGGGGAAAGCCGCTCTTATCACGGCCGCA
TTCGAAACCTATGATTCGGTTTGTCGCGAGCGTGCGCAGTGGCTGGTGGAAAGTA
GTCGCGTTATCGGTAATCTGTATGAGTGGCAGGATAAGGAGGTAGGGTCGGATGC
TTCCAGGTGCCACGATGAGGTGTATTGGCGCTCTCATCGCATTTGGGACTATGATA
TTGATGCGATGATGAGAGAGACAGCTGAGGTGTTTGAGGCGCAGGTAGCTGGGGT
GGCGAGAAAT

AfoD Protein Sequence

MADHEQEQEPLSIAIIGGGIIGLMTALGLLHRNIGKVTIYERASAWPDIGAAFAFTGIARE
CMQRLDPAILSALSKVAQRNPHDKVRYWDGFHPKSKEEAQDPEKSVLFEIEEKNMAY
WACLRGVFHAEMARLLPERVVRFGKRLVAYEDGGDQKVVLRFEDGEVEEADIVIACD
GVHSTARRVLLGAEHPAANARYSRKAVYRALVPMPAAIDALGTEKAHVQIAHCGPDAH
IVSFPVNNAQIYNVFLFTHDSNEWTHGHTMTVPSSKEEILSAVENWGPHIKELASLFPE
QLSKYAIFDQADHPLPYYAAGRVALAGDAAHASSPFHGAGACMGVEDALVLAELLEKV

QNGSAFKEKKSNIELALKTYSDVRIERSQWLVKSSREMGDLYEWRYEDIGGDGVKCK
AEWERRSRVIWDFDVQGMVDQAREAYERAVVKV

Codon-Optimized *fdmo1* Sequence

ATGCCGAGCTATAACAAAGATACCGAAAGCGTGGAAGTGGCGGTGATTGGCGGCG
GCATTGTGGGCCTGGTGCTGGCGGCGGGCCTGACCCGCCGCCAGATTAAAGTGA
AAGTGTATGAACAGAGCCAGGGCTTTCGCGATATTGGCGCGGGCATTGGCTTTAA
CGGCGCGGCGAAAGCGTGCATGCAGATGATTGATCCGGGCGTGATTACCGCGCT
GCATCGCGGCGGCGGCGTGGCGGTGAGCGCGGCGGATGAAGATGATCCGCATG
ATTATCTGCGCTGGATTGATGGCTTTGATCGCGGCAACGTGCAGCATCTGCATGAT
CAGAAACTGTATTGCAAAGTGGATGCGGGCTATAAAAGCATTGAAGGCACCCGCC
GCGATCGCTTTCTGGAAGAACTGGCGAAAGATCTGCCGGAAGGCATGGTGGAATT
TAAAAAACGCCTGCGCACCGTGGAAGAAGGCGGCGATGATTGCAAACTGCAGCTG
CATTTTGAAGATGGCACCATTGCGGAAGCGGATGCGCGCTGCGATGGCATTAAAA
GCCGCATTCGCGAAATTGTGCTGAGCGAAGCGAGCGTGGCGAGCAAACCGAGCT
ATACCCATGTGAACTTTTATAGCAGCCTGATTCCGATGAACAAAGCGGTGGATATT
CTGGGCAAATTTAAAGCGAGCGTGTTTCATAACCATATTGGCCCGGGCGCGAACG
TGCTGCATTATCCGGTGGCGAACGGCACCCTGTGCAACGTGAGCGCGTTTGTGCA
TGATGCGAACGAATGGCCGGCGGAAAAAAGCCCGACCAGCATTGGCTTTCGCAAA
CATATTCAGGAAAAACTGGTGGGCTGGAGCCCGGTGGTGCGCGGCCTGATTGATC
TGTTTCCGGATACCCTGCCGGTGTGGGCGGTGTTTGATCTGTGGGAACATCCGAT
GCCGTATTATAACCGCGGCCGCATTTGCGTGGCGGGCGATGCGGCGCATGCGAG
CAGCCCGCATCATGGCGCGGGCGCGGGCATGGGCATTGAAGATGCGCTGTGCCT
GAGCGTGCTGCTGGATGAAGTGAGCAGCAGCATTCGCCTGGAAGGCGCGAGCCG
CCGCGATGCGATTCCGGTGGCGTTTCAGGTGTATGATAGCATTCGCCGCCGCCGC
AGCCAGTGGCTGGTGAACAGCAGCCGCCGCCTGTGCGATCTGCAGCAGCATCAT
GATTGGGCGGATCCGGCGAAACTGGTGAAAGCGGAAACCTGCTTTGAAGAAATTA
CCGATCGCACCTATAAAATTTGGAACTTTGATAGCAACGGCATGATTAAAGAAAGC
ATTGAAAAATATGGCCGCGCGATTAACAGCCTGCGCCGCAACGGCCTGGCGACCA
ACACCGATTGCAAAGGCAACGGCCATATGAACGGCGTGCGCGCG

Codon-Optimized *fdmo2* Sequence

ATGGCGAGCACCGAACCGCAGGCGGATAGCGTGGATGTGGTGATTGTGGGCGGC
GGCATTATTGGCCTGGTGCTGACCGTGGGCCTGCTGCGCGTGGGCGTGAAAGTG
AAAGTGTATGAACAGGCGCAGGGCTTTCGCGAAATTGGCGCGGGCATTGCGTTTA
CCGCGAACGCGATTCGCTGCATGAACCTGATTGATCCGGCGATTCCGGTGGCGCT
GCGCAGCAGCGGCAGCGTGGCGACCAGCAACGGCGGCGATGAAGATCCGAACG
ATTATCTGCGCTGGATTGATGGCTATGATCGCCAGCGCGATGATCCGAGCCTGCA
GCAGCTGTTTTTTAAACTGAACGCGGGCTATCGCGGCTTTGAAGGCTGCCGCCGC
GATCAGTTTCTGGAAGCGCTGGTGAAAGTGATTCCGCCGGGCGTGATTGAACTGA
AAAAACGCCTGGAAACCGTGCATGATAACGGCAGCGAAAACAAACTGCTGCTGAC
CTTTCAGGATGGCACCACCGCGGAAGCGGATGCGGTGATTGGCTGCGATGGCGT
GAAAAGCACCCTGCGCCGCATTATGTTTGGCGATGATCATCCGGCGAGCCGCCCG
CGCTATAGCCATTGCGTGGCGTATCGCACCCTGATTCCGATGGATAAAGCGGTGA
GCGCGCTGGGCGCGTATAAAGCGACCAACCAGCATAACCATGTGGGCCCGAACG
CGAACATTCTGCATTATCCGGTGGCGAACAACACCATGATTAACGCGGTGGCGTTT

ATTCGCGATCCGAACGAATGGACCGATGAAAAAACCGTGGCGGAAGGCACCCGC
GATGATGTGAAAGCGGCGGTGCGCGGCTGGAGCCAGCCGGTGCTGAACCTGGTG
GATTGCTTTCCGGATACCCTGAGCAAATGGGGCATTTTTGATCTGTGGGAATTTCC
GGTGCCGAGCTATAACGTGGGCCGCCTGAGCCTGGCGGGCGATGCGGCGCATG
CGAGCAGCCCGCATCATGGCGCGGGCGCGTGCATGGGCATTGAAGATGCGCTGT
GCCTGACCACCCTGATGGAACAGGTGGTGGTGGAAGCGCAGAAAAGCCCGGGCG
ATAAAGGCCGCGCGCTGATTGCGGCGCTGGATACCTATAGCGCGGTGCGCCAGA
CCCGCAGCCAGTGGGTGGTGAACAGCAGCCGCCGCGTGTGCGATCTGCATCAGC
AGCAGGAATGGGCGGATGCGACCAAACTGATTAAAGCGCAGACCTGCTTTGAAGA
AGTGAAAGATCGCAGCCTGAAAATTTGGCATTTTGATTATGAACGCATGGTGCGCG
ATAGCCTGCAGGGCTATAAACAGCGCCGCGCGCCGATTAACGGCGCGACCAAAG
ATAAAAACCTGTAT

Codon-Optimized *fdmo3* Sequence
ATGATTGAAGCGCGCGCGATTGAAGTGGCGATTATTGGCGGCGGCATTACCGGCC
TGACCCTGGCGCTGGGCCTGCAGAAACGCAACACCAACTTTCATATTTATGAACGC
GCGCAGAGCCTGCGCGAAATTGGCGCGGGCATTGGCTTTACCCCGAACGCGGAA
CGCGCGATGCTGGCGCTGGATCCGCGCATTCATGAAGCGTTTAAAAGCGTGGCGA
GCAAAAACGCGAGCGATTGGTTTCAGTGGGTGGATGGCTTTAGCGGCGTGAACAA
CGATAAAGATACCGTGAAAGAAGATCTGCTGTTTAACATGTATCTGGGCGAACGCG
GCTTTGAAGGCTGCCATCGCGCGCAGTTTCTGAAAGAACTGGTGAACCATCTGCC
GCAGGGCTGCGTGACCTTTGGCGCGTGCCTGGATACCATTATTGATCAGGGCGAA
AACGAACGCATTCTGCTGAAATTTCATAACGGCACCATTGCGGAAGCGGATCTGGT
GATTGGCTGCGATGGCATTCGCAGCCGCGTGCGCCAGCTGATTCTGGGCGAAAA
CAACCCGGCGAGCTATCCGGCGTATACCCATAAAAAAGCGTATCGCGGCCTGATT
CCGATGGAAAAAGCGCTGCCGGCGCTGGGCGAAAGCAAAGTGAACACCCGCCTG
ATGCATCTGGGCCCGGATGCGCATACCCTGACCTTTCCGGTGGCGGGCGGCAAA
CTGATGAACGTGGTGGCGTTTGTGACCGATCCGGGCGAATGGCCGTATACCGAAA
AACTGAGCGCGCCGGCGGAAAAAAAAAGCGCGATTGAAGGCTTTAGCAAATTTGG
CGGCGCGGTGCGCACCATTATGAACCTGCTGCCGGAAGATCTGGATGAATGGGC
GATTTTTGATACCTATGATCATCCGGCGAGCACCTATTATCATGGCCGCATTTGCAT
TGCGGGCGATGCGGCGCATGCGAGCAGCCCGCATCATGGCGCGGGCGCGGGCG
CGGGCATTGAAGATGTGACCGTGCTGGCGACCGTGATTGAAGTGGCGCAGACCA
CCCTGCTGGAAAGCCCGGATAAAAGCCGCAGCGGCGTGCTGAACGCGGCGCTGG
CGACCTATAACGCGGTGCGCCTGGAACGCAGCCAGTGGCTGGTGGAAAGCAGCC
GCATTCTGGGCGAAATTTATGAATGGCAGTATAAACCGACCGGCCGCGATAAAAAA
AAATGCGAAGAAGAAGTGTATTGGCGCAGCCATAAAATTTGGGATTATGATATTGG
CCAGATGCTGCAGGAAACCACCGAATATTATAAACAGCGCGTGGGCGCG

Codon-Optimized *fdmo4* Sequence
ATGGATACCAACAAATTTGAAATTGCGATTATTGGCGCGGGCATTACCGGCATTAC
CCTGGCGCTGGGCCTGCTGAGCCGCGGCATTCCGCCGCGCGATTTTCATGAAATT
GGCGCGGGCATTGGCTTTACCCCGAACGCGGAATGGGCGATGAAAGTGGTGGAT
CCGCGCATTCATGCGGCGTTTAAACGCGTGGCGACCCCGAACGCGAGCGATTGG

TTTCAGTGGGTGGATGCGTTTAACGAAACCGGCGAACGCGGCTTTGAAGGCTGCC
ATCGCGCGCAGCTGCTGGGCGAACTGGCGCGCCTGCTGCCGGAAGGCATTGTGA
CCTTTTATAAAGCGCTGGATACCCTGGAACCGGCGGCGGATAACCGCCTGGGCCA
GCTGCTGCGCTTTCAGGATGGCACCACCGTGACCGCGCATGCGGTGATTGGCTG
CGATGGCATTCGCAGCCGCGTGCGCCAGATTCTGTTTGGCGAAGATCATCCGGCG
GCGAGCGCGCATTATAGCCATAAATATGCGGCGCGCGGCCTGATTCCGATGGATC
GCGCGCGCGAAGCGCTGGGCGATGCGAAAGTGGCGACCCGCTTTATGCATCTGG
GCCCGGATGCGCATGCGCTGACCTTTCCGATTGCGCATGGCAGCCTGCTGAACGT
GGTGGCGTTTGTGACCGATCCGAACCCGTGGCCGTATGCGGATCGCTGGACCGC
GCAGCGCAACGAAACCGATGTGGCGGCGGCGTTTAGCCGCTTTGGCCCGACCAT
GCGCACCATTATTGATCTGCTGCCGGATCCGATTGATCAGTGGGCGGTGTTTGATA
CCTATGATCATCCGCCGAACACCTATAGCCGCGGCCCGGTGTGCATTGCGGGCGA
TGCGGCGCATGCGGCGGCGCCGCATCATGGCGCGGGCGCGGGCTGCGCGGTGG
AAGATGTGGCGGTGCTGTGCGCGGTGCTGGATCTGGCGGCGAAACGCGTGGATG
CGACCAAATGCGATCCGAAAGGCAAAGCGGCGCTGATTACCACCGCGTTTGAAAC
CTATGATGCGGTGCGCCGCGAACGCGCGCAGTGGCTGGTGGAAACCAGCCGCAT
TATTGGCAACTTTTATGAATGGCAGGATAACGAAGTGGGCCCGGATGCGAGCATTT
GCCATGATGAAGTGTATTGGCGCAGCCATCGCATTTGGGATTATGATATTGATACC
ATGATGCGCGAAACCGCGAAAGTGTTTGAAGTGCGCGTGGCGGAACTGACCAAAA
AC

Codon-Optimized *fdmo5* Sequence
ATGGCGAGCAACAACAAAACCACCAACCCGAGCATTGAAGTGGCGGTGGTGGGC
GGCGGCGTGATTGGCGTGATGACCGCGCTGGGCCTGATTCGCCGCGGCATTAAA
GTGACCATTTATGAACGCAGCAGCAACTGGCATGAAATTAGCGCGGGCTTTGCGT
TTACCGGCGTGGCGCGCGAATGCATGCAGCGCCTGGATCCGGGCATTCTGGATG
TGCTGAGCCGCATTAGCCAGAAAACCGATCCGAACGATAGCAGCACCACCTATTG
GAACGCGTATCATCCGCAGACCAAACAGGATGCGGAAGATGAAAGCACCAGCCTG
CTGTTTCAGCTGCCGGGCAACAAACTGGCGTTTTGGGGCTGCGTGCGCAGCCAGT
TTCTGCTGGGCATGGTGGCGCTGCTGCCGGATGATGTGGCGCGCTTTGGCAAACA
GCTGGTGAGCTATGATGATGGCGATGCGAACGATAAAGTGGTGCTGCATTTTGCG
GATGGCAGCACCGCGGAAGCGGATGTGGTGCTGGGCTGCGATGGCATTCATAGC
ACCACCCGCAAAACCCTGCTGGGCGCGCATCATCCGGCGACCCGCCCGAGCTAT
ACCCATACCGTGGCGTATCGCACCATGGTGCCGATTGATGCGGGCATTGCGGCGC
TGGGCGAAGATAAAGCGCGCCGCGCGTGCATGCATTGCGGCCCGAACGCGAACA
TGATGAGCTATCCGGTGATGAACGGCACCCTGCTGAACGTGGCGTTTTTTGCGCA
TGAAAGCAGCGAATTTCCGGATCCGGAAAAAATGACCGCGCCGGGCACCCGCGA
AGAACTGGAACGCGTGGTGGTGGGCTGGGGCCCGCATCTGGTGGAACTGACCAA
ACTGTTTCCGGATAACATGGTGAAATGGGGCATTTTTGATATGGATGAAAACCCGG
CGCCGACCTATGCGCGCGGCTGCGTGTGCCTGGCGGGCGATGCGGCGCATGCG
AGCAGCCCGTTTCAGGGCGTGGGCGCGTGCATTGGCGTGGAAGATGCGCTGGTG
CTGTGCGAAGCGCTGGCGACCGTGCAGGCGGGCGGCAACAGCGGCAGCGATGA
TGGCAACCATACCCATAGCCAGCGCGAAGTGATTGAACAGGCGCTGCAGGCGTAT

AGCCAGGCGCGCATTGATCGCGGCCAGTGGGTGGTGCGCAGCAGCCGCGAACTG
GGCCAGATTTATCAGTGGCGCTATGGCCCGACCGGCCGCGATGCGGAACGCAGC
AAACTGAAACTGGAACGCGCGAGCCGCACCGTGTGGGATTATGATGTGGATAAAA
TTGTGACCGAAATTCGCGCGGTGGTGGCG

Codon-Optimized *fdmo6* Sequence

ATGACCGTGGCGGATCGCGCGCCGCTGGATGTGGCGATTATTGGCGGCGGCATT
ATTGGCATTATGACCGCGCTGGGCCTGCTGCATCGCGGCTTTCGCGTGACCGTGT
ATGAACGCGCGGCGAGCTGGCCGGAAATTGGCGCGGCGTTTGCGTTTACCGGCG
TGGCGCGCCAGTGCATGGAACGCCTGGATCCGCGCGTGCTGGAAAGCCTGGCGC
GCGTGGCGCAGCGCAGCCCGCATGAAAAAGTGCGCTATTGGGATGGCTTTCATCC
GCGCACCAAAGAAGCGGCGCAGGAAGAAAGCGCGGTGCTGTTTGAAATTCTGGAA
AAACATATGGCGTATTGGGCGTGCATTCGCGGCCATTTTCTGCTGGATATGGCGG
CGCAGCTGCCGGATGGCGTGGTGCAGTTTGGCAAACGCCTGGTGGATTATAACGA
TGATGAAGCGAACGAAAAAGTGGTGCTGTGCTTTGCGGATGGCAGCACCGCGGAA
AGCGATGTGGTGATTGCGTGCGATGGCATTCATAGCGCGACCCGCAAAGTGCTGC
TGGGCGTGGATCATCCGGCGGCGAACGCGAGCTATAGCCGCAAAAGCATGTATC
GCGCGATGGTGCCGATGGCGGATGCGGTGAGCGCGCTGGGCACCGAAAAAGCG
CATGTGCAGATTGCGCATCTGGGCCCGGATGCGCATGTGGTGAGCTTTCCGGTGA
ACAACGGCCAGGTGTATAACGTGTTTCTGTTTCTGCATGATCCGAACGAATGGGAT
CATGGCCATACCATGACCGTGCCGAGCAGCCGCAGCGAAGTGATGGATGCGATTC
AGGGCTGGGGCCCGCATATTAAAGAAATTGTGAGCTGCTTTCCGGAAACCGTGAG
CAAATATGCGATTTTTGATCAGGCGGATAACCCGCTGCCGTATTATGCGAGCGGC
CGCGTGTGCCTGGCGGGCGATGCGGCGCATGCGAGCAGCCCGTTTCATGGCGCG
GGCGCGTGCATGGGCGTGGAAGATGCGCTGGTGCTGGCGGAACTGCTGGGCCTG
GTGGATGCGGGCCCGGTGGCGGCGCGCCAGCGCAACATTAAAGCGGCGCTGCA
GACCTATAGCAGCGTGCGCATTGAACGCAGCCAGTGGCTGGTGCAGAGCAGCCG
CGATATGGGCGATCTGTATGAATGGCGCTATCCGCCGACCGGCGAAGATGGCGC
GAAATGCAAAGCGGAATTTGAACGCCGCAGCAAAGTGATTTGGGATTTTGATGTGG
ATGGCATGGTGGCGGGCGCGAAAAAAAAATATGAACATAGCATGGAAGCG

Codon-Optimized *fdmo7* Sequence

ATGGAAGCGCCGAACAACCATCCGAACGGCATTAACGTGATTAACGGCCATAAAG
CGAAAAGCCTGGAAGTGGCGATTGTGGGCGGCGGCCTGACCGGCCTGGCGCTG
GCGGTGGGCCTGCTGCGCCGCAACATTAACTTTACCATTTATGAACGCGCGGCGA
GCTTTGGCGAACTGGGCGTGGGCATTCATTTTACCCCGAACGCGGAACGCGCGAT
GGAAGCGCTGGATCCGCGCGTGCTGCAGAGCTATGTGGATGTGGCGACCAACGC
GGAAGGCGGCTTTCTGAGCTTTGTGGATGGCGCGAGCGGCGATGATGGCCTGCT
GTTTCAGCTGCGCATGGGCAAAGGCTATAAAGCGGCGCGCCGCTGCGATTTTGTG
AGCCAGCTGGTGAAACATATTCCGCAGGAACGCGTGCAGCATCTGAAATGGCTGC
AGAGCGTGGAAGAAGATGGCGAAGGCCGCGCGGTGCTGACCTTTCGCGATGGCA
GCACCGCGGAAGCGGATGTGGTGGTGGGCTGCGATGGCATTCGCAGCCAGGTGC
GCAGCGCGATGTTTGGCAGCGGCCCGAGCGCGCCGCGCGCGCAGTATGCGCAT

CAGCTGGCGTTTCGCGGCCTGGTGCCGATGGCGAAAGTGGAAGAAGCGCTGGGC
AGCGGCAAAACCAGCCGCGCGATTGGCTATCTGGGCCCGGGCGGCTTTGTGCTG
AGCGTGCCGCTGGCGGGCATTAACATGATGCATCTGGAAGTGTTTGTGATGGATC
CGCTGGATTGGAGCGATACCCGCAGCAAAAGCGAAAAAGGCAACGATGAAGATGA
TGTGAAACGCTATGTGCTGCCGGCGACCCGCGCGGAAGCGGAAAAAGCGTTTAC
CGAATTTAACCCGACCGTGCGCAGCCTGATTAGCCTGCTGCCGGAAACCCTGGGC
AAATGGGCGATTTTTGATATGCTGGATAGCCCGGCGCCGAGCTATGCGCTGGGCC
GCATGTGCCTGGCGGGCGATGCGGCGCATGCGAGCACCCCGAACCAGGGCGGC
GGCGCGGGCGCGGGCATGGAAGATAGCCTGGTGCTGGCGGAAATTCTGGCGGC
GCTGGCGGATCGCGAAAACAGCGGCGCGCCGGTGGGCCTGAGCGAAATTAGCGA
AGGCCTGAAAGTGTATAGCGAAGCGCGCTATGAACGCGCGCAGTGGCTGGTGCA
GAGCAGCCGCCGCGTGGCGCAGCTGTTTACCCGCAAAAGCGCGGAACAGGAAGA
ACCGATTAGCCGCGAAATTCTGGAACGCAGCCATCAGCTGTGGGATCATGATGTG
GATGCGATGGTGGCGGATGCGCTGGGCAAACTGAAAGCGAAACTGAGCGAAAAA
AAA

Protein overexpression and purification: Plasmids containing the genes of interest were transformed using standard heat-shock protocols into chemically competent *E. coli* into BL21(DE3) cells. Overexpression of AfoD was achieved in 500 mL 4% glycerol (v/v) Terrific Broth (TB) in 2.8 L flasks. 500 mL portions of media were inoculated with 5 mL overnight culture prepared from a single colony in Luria Broth (LB) and 100 μg/mL ampicillin (Gold Biotechnology). Cultures were grown at 37 °C and 250 rpm until the optical density at 600 nm reached 0.8. The cultures were then cooled to 18 °C for 1 h and protein expression was induced with 0.1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG, Gold Biotechnology). Expression continued at 20 °C overnight (approx. 18 h) at 200 rpm. The typical yield for one 500 mL culture was ~15 g cell pellet. Overexpression of AzaH followed the same protocol as described above, except 1 L cultures were grown in 2.8 L flasks and kanamycin was used at 50 μg/mL (Gold Biotechnology) in place of ampicillin. The typical yield for one 1 L culture was ~30 g cell pellet.

General purification procedure: 25-30 g of cell pellet was resuspended in 100 mL of lysis buffer containing 50 mM Tris HCl pH 7.4, 300 mM NaCl, 10 mM imidazole, and 10% glycerol. Protease inhibitors were added to lysis buffer of AzaH only and consisted of 1 mM phenylmethane sulfonyl fluoride (v), 0.1 mg/mL benzamidine HCl, 0.5 mg/mL leupeptin, and 0.5 mg/mL pepstatin. Approximatley 1 mg/mL lysozyme was added to resuspended cells that were then incubated on a rocker at 4 °C for 30 min. Cells were lysed by passing the total cell lysate through an Avestin pressure homogenizer at 15000 psi. The total lysate was centrifuged at 40,000 x g for 30 min and the supernatant was filtered through a 0.45 μm filter. The crude cell lysate was loaded onto a 5 mL HisTrap HP column (General Electric) on an ÄKTA Pure FPLC system (General Electric) at a flow rate of 2.5 mL/min. Buffer A = the lysis buffer listed above, and Buffer B = 50 mM Tris HCl pH 7.4, 300 mM NaCl, 10% glycerol, and 400 mM imidazole. The column was washed with 25 mM imidazole (6.3% Buffer B) for 6 CV and eluted in a gradient to 100% Buffer B over 8 CV. Fractions containing AfoD or AzaH were visibly yellow and pooled for desalting on a PD10 desalting column. Average yields: 100 mg from 1 L AfoD, 20 mg from 1 L AzaH. Molecular weights including 6xHis-tags for each protein were estimated by the ProtParam tool on the Expasy server to be 49.0 kDa for AfoD and 47.6 kDa for AzaH. These molecular weights are consistent with the mass of proteins bands observed by SDS-PAGE analysis (Figure 2.3). The purified proteins were aliquoted into 0.6 mL tubes and frozen in liquid nitrogen before long-term storage at -80 °C.

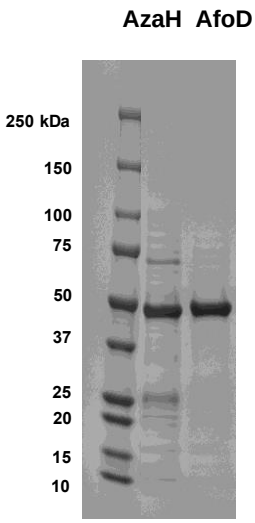


Figure 2.3. Purified AzaH and AfoD. Approximately 5 µL of 1.25 µM each protein was loaded onto an MiniPROTEAN TGX Precast 4-15% SDS-PAGE gel (Bio-Rad). The gel was stained with Quick Coomassie stain (Anatrace) and visualized with the Azure Gel Imaging System. The relative apparent masses are consistent with the predicted estimates.

Determination of flavin incorporation and extinction coefficients: Samples of each protein were diluted to 10 µM in 1 mL using dialysis buffer for UV-vis analysis using a disposable poly(methyl 2-methylpropenoate) cuvette. The absorbance spectrum for each protein was taken from 300 nm to 700 nm in 2 nm increments (blue traces in Figure 2.4). A 20 µL aliquot of fresh 10% sodium dodecyl sulfate (w/v) was added to each 1 mL solution and mixed. Samples were incubated at room temperature for 10 min before reading the absorbance spectra again under the same conditions (red traces in Figure 2.4). The absorbance at 450 nm for the denatured enzymes and the extinction coefficient of free FAD (11300 M^{-1} cm^{-1}) was used to calculate the concentration of FAD in each protein sample using Beer's law. The typical FAD incorporation was 82% for AzaH, 81% for AfoD. Extinction coefficients were calculated using the concentrations of free flavin obtained and the absorbance at 450 nm of the native enzymes. At 450 nm, the extinction coefficients of the proteins are 17490 M^{-1} cm^{-1} for AzaH, 6,870 M^{-1} cm^{-1} for AfoD.

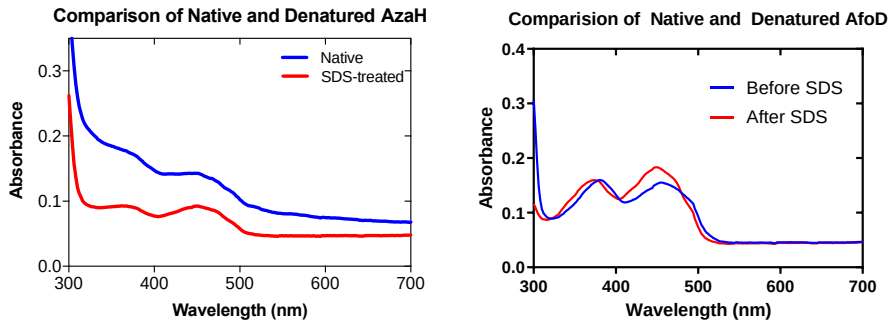


Figure 2.4. Native enzyme absorbance spectra compared to denatured enzyme absorbance spectra exposing free FAD to solution.

III. Generation of AfoD Y118F variant

General Considerations

E. coli cloning strains DH5α (Invitrogen) were used for DNA propagation. Phusion HF polymerase was purchased from New England BioLabs. All primers were purchased from Integrated DNA Technologies (IDT) ddH2O was sourced from a MilliQ Biocel water purification unit from Millipore.

Site-directed mutagenesis

Table 2.1. Primer Sequences

Variant	Plasmid ID	Oligo sequence
AfoD Y118F	AfoD_Y118F_P1	GAAAAGAACATGGCATTTTGGGCATGTC

The AfoD(Y118F) substitution was generated by site-directed mutagenesis on pET151-afoD(WT). 25 µL PCR reaction mixture contained 5 µL of 5X Phusion HF buffer, 1 ng/µL WT parent plasmid, 0.5 µM of primer, 200 µM dNTPs, 0.5 U µL-1 Phusion HF.

Amplification was accomplished with the following PCR protocol: 95 °C for 0:30 s, (95 °C 0:30 95 °C for 0:30 s (-0.5 °C/cycle), 72 °C 0:30/kb) for 12 cycles, (95 °C for 0:30 s, 65 °C for 0:30 s, 72 °C 0:30/kb) for 20 cycles with a final extension of 72 °C for 10:00 min. This was followed by a 10 µL digestion containing 1 µL of NEB CutSmart buffer, 8 µL of PCR mixture and 20 units of DpnI. The reaction was incubated at 37 °C for 3 h and transformed into chemically competent E. coli DH5α cells.

Protein Expression and Purification

Protein overexpression: AfoD(Y118F) plasmid was transformed into E. coli strain BL21(DE3). 500 mL of Terrific Broth (TB) containing 100 µg mL-1 ampicillin was inoculated with 5 mL overnight culture prepared from a single colony in Luria Broth (LB) and 100 µg mL-1 ampicillin. The culture was grown at 37 °C and 250 rpm for 4 h. The culture was then cooled to 20 °C for 1 h at 200 rpm, and protein expression was induced with 0.1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and expressed at 20 °C for 18 h at 200 rpm. After overnight expression, cultures were centrifuged at 13,881 x g for 30 min. Cell pellets from overexpression were stored at -80 °C for long-term storage.

General purification procedure: Cell pellets from overexpression were resuspended in 40 mL of lysis buffer (50 mM Tris:HCl pH 7.8, 300 NaCl, 10 mM imidazole, and 10% (v/v) glycerol) with 0.1 mg mL-1 lysozyme, 0.05 mg mL-1 DNase, and 0.1 mM flavin adenine dinucleotide (FAD), incubated on a rocker at 4 °C for 45 min, and lysed by sonication. Insoluble material was removed by centrifugation (46,413 x g for 30 min). The cell pellet was resuspended in 40 mL lysis buffer (50 mM Tris:HCl pH 7.8, 300 NaCl, 10 mM

imidazole, 10% (v/v) glycerol) with 0.1 mg mL-1 lysozyme, 0.05 mg mL-1 DNase, and 0.1 mM FAD, incubated on a rocker at 4 °C for 45 min, lysed by sonication and cleared by centrifugation (46,413 x g for 30 min). The supernatant was incubated with Ni-NTA on a rocker for 2 h at 4 °C, followed by purification by gravity using a 25-50 mM gradient with increments of 5 mM imidazole. Protein was eluted with 100% elution buffer (50 Mm Tris:HCl pH 7.8, 300 mM NaCl, 400 mM imidazole, 10% (v/v) glycerol). Concentrated protein was desalted over a PD-10 desalting column (GE Healthcare). The protein was concentrated further using a 30 kDa centrifugal concentrator at 4,000 x g, 4 °C. Concentrated protein was divided into 100 µL aliquots, frozen in liquid nitrogen and stored at -80 °C.

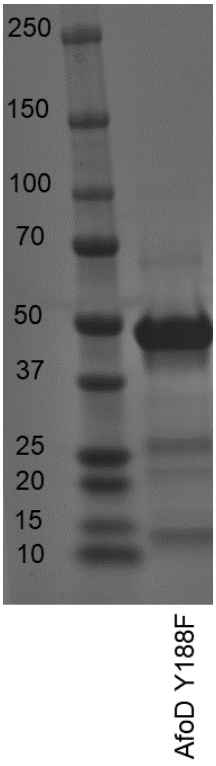


Figure 2.5. AfoD(Y118F) SDS-PAGE gel. The gel was stained with Quick Coomassie stain (Anatrace) and visualized with the Azure Gel Imaging System

IV. Sequence Similarity Network (SSN)

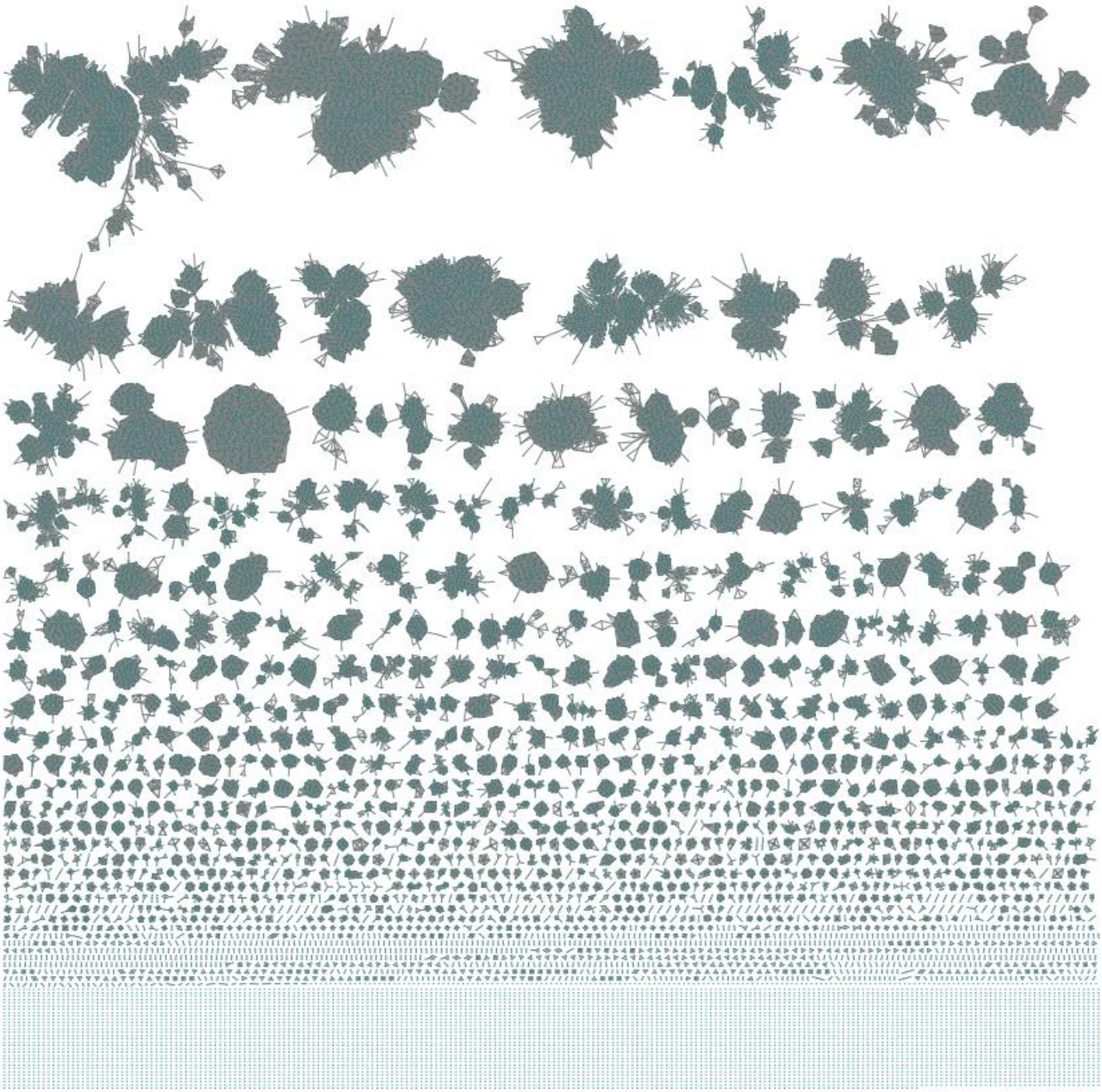


Figure 2.6. SSN of flavin-dependent monooxygenases (pfam 01494) created using web tools originating from the Enzyme Function Initiative (EFI). [30-32]

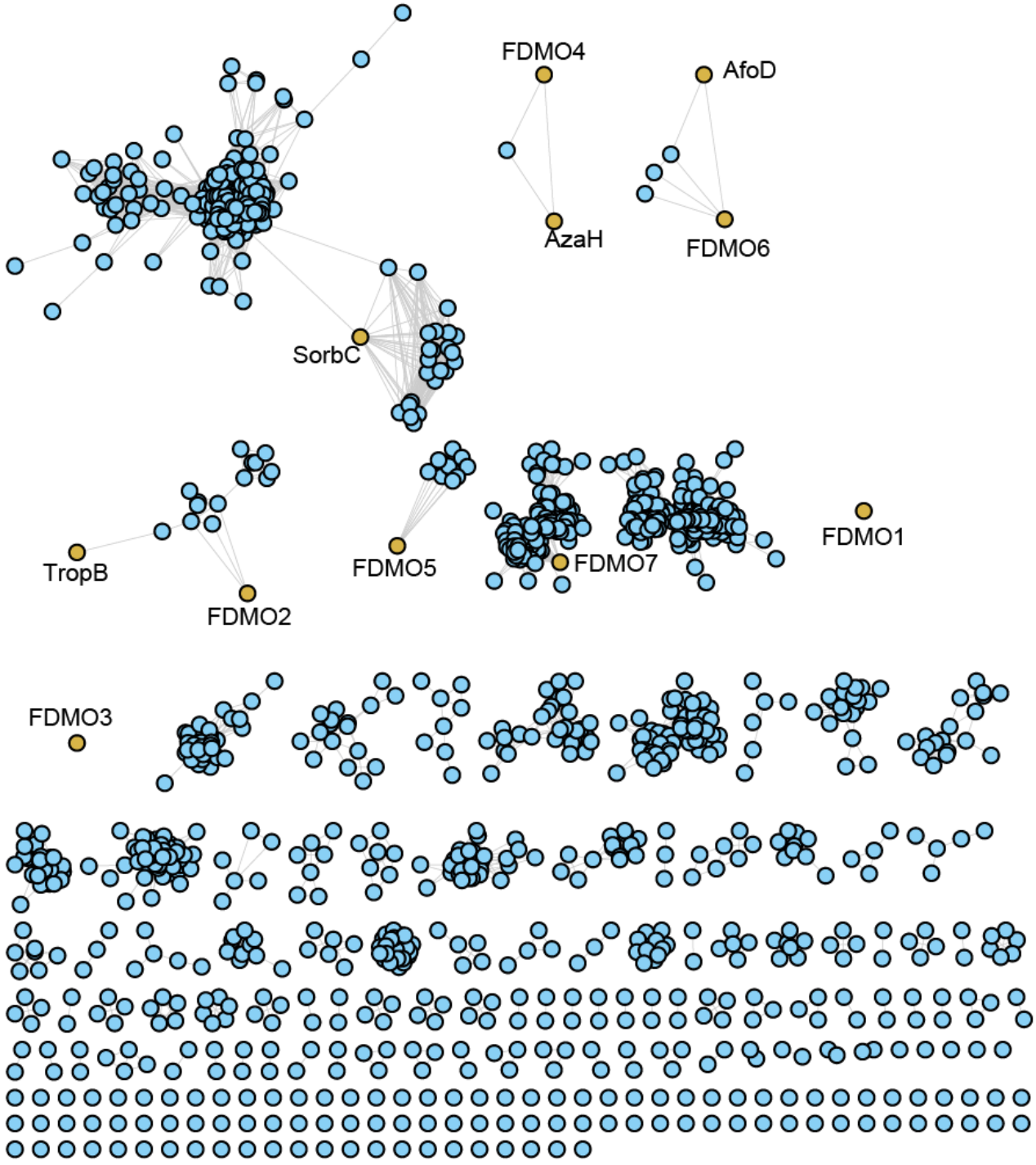


Figure 2.7. Refined sequence similarity network with E value of 150.[30-32]

V. FDMO Sequence Alignment

Figure 2.8. AfoD cluster sequence alignment with other sequences within its cluster Y118 and F237 are highlighted. Alignment is colored by conservation.

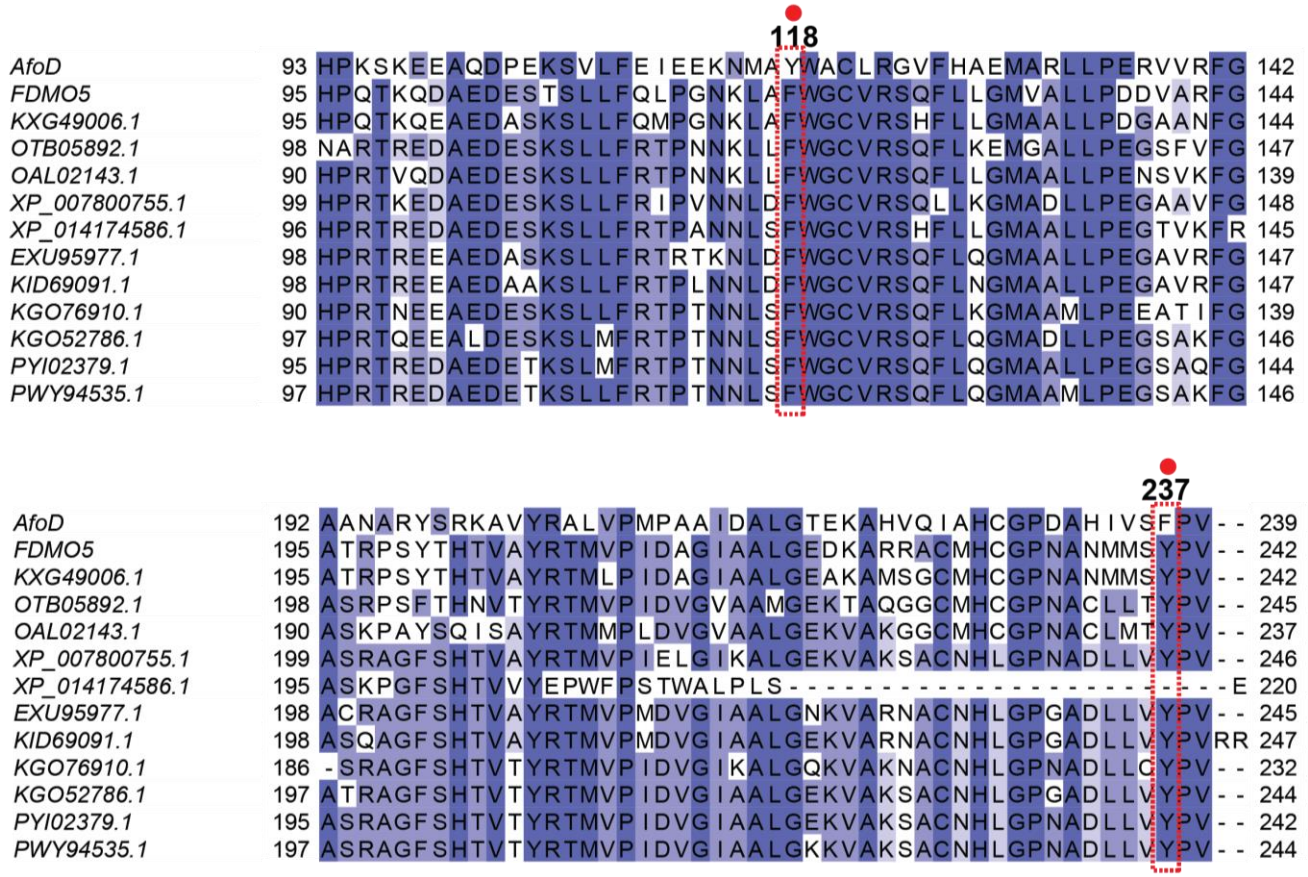


Figure 2.9. FDMO5 cluster sequence alignment with other sequences within its cluster Y118 and F237 are highlighted (From AfoD). Alignment is colored by conservation.

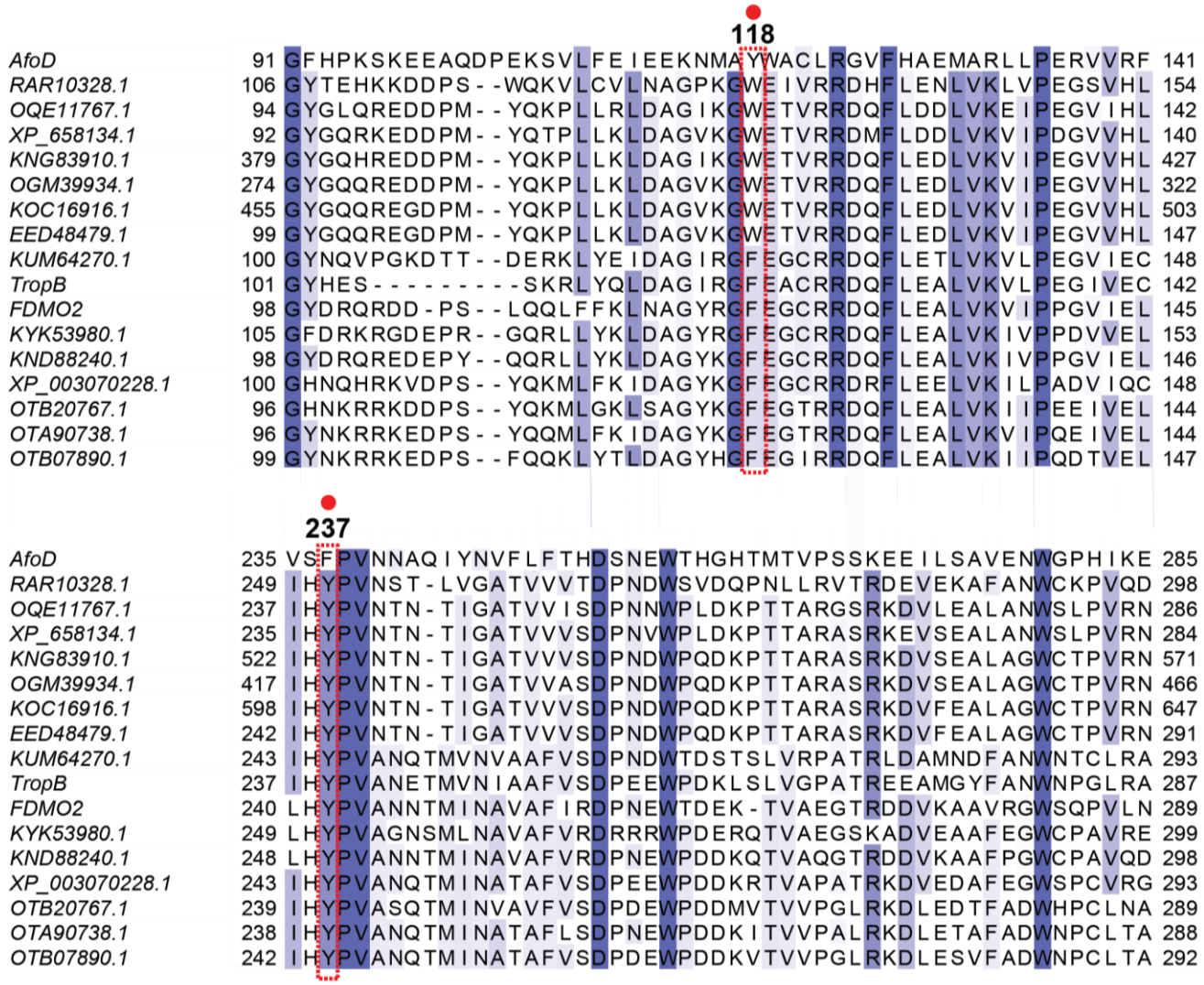


Figure 2.10. TropB cluster sequence alignment with other sequences within its cluster Y118 and F237 are highlighted (From AfoD). Alignment is colored by conservation.

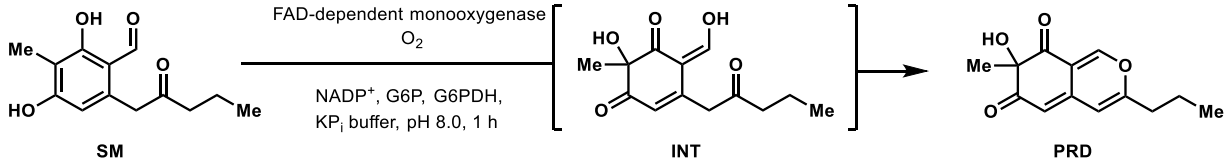
VI. Biocatalytic Reactions

Stock solutions: Stock solutions of each substrate (50 mM) were prepared by dissolving the substrate in DMSO (analytical grade). Stock solutions of $NADP^+$ (100 mM) and glucose-6-phosphate (G6P, 500mM) were stored at -20 °C. Aliquots of each flavin-dependent enzyme and glucose-6-phosphate dehydrogenase (G6PDH, 100 U/mL) were stored at -80 °C. Analytical-scale reactions: Each reaction contained 25 µL 100 mM potassium phosphate buffer, pH 8.0, 2.5 mM substrate (2.5 µL of a 50 mM stock solution in DMSO), 5-20 µM flavin-dependent monooxygenase, 5 mM G6P (0.5 µL, 500 mM), 1 mM NADP+ (0.5 µL, 100 mM), 1 U/mL G6P-DH (0.5 µL, 100 U/mL), and Milli-Q water to a final volume of 50 µL. The reaction was carried out at 30 °C for 1 h and quenched by addition of 75 µL acetonitrile with 25 mM pentamethylbenzene as an internal standard. Precipitated biomolecules were pelleted by centrifugation (16,000 x g, 12 min). The supernatant was analyzed by UPLC-DAD and conversion obtained by comparison to calibration curves of each substrate. The subsequent liquid chromatography PDA spectrometry (UPLC) analysis was performed on a Waters Aquity H-Class UPLC-PDA using a Phenomenex Kinetex 1.7 µm C18, 2.1x150 mm column under the following conditions: Method A: mobile phase (A = deionized water + 0.1% formic acid, B = acetonitrile + 0.1% formic acid), 5% to 100% B over 1.5 min, 100% B for 1.0 min; flow rate, 0.5 mL/min; Method B: mobile phase (A = deionized water + 0.1% formic acid, B = acetonitrile + 0.1% formic acid), 5% to 100% B over 2 min, 100% B for 1 min; flow rate, 0.5 mL/min. Based on calibration curves of the starting materials, the percent conversion of the substrate to dearomatized product was calculated with AUCsubstrate/AUCinternal standard at 270 nm. All reactions were performed and analyzed in triplicate.

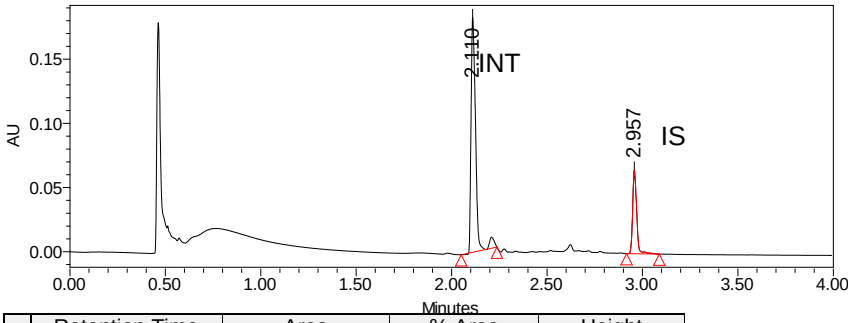
General procedure for *in vitro* preparative-scale reactions: Preparative–scale enzymatic reactions were conducted on 20 mg of each substrate under the following conditions: 5-20 µM flavin-dependent monooxygenase, 2.5 mM substrate, 1 mM NADP+, 1 U/mL G6PDH, and 5 mM G6P for NADPH generation in reaction buffer (50 mM potassium phosphate buffer, pH 8.0). The reaction mixture was added to a 50 mL Erlenmeyer flask and incubated at 30 °C with 100 rpm shaking. After 2 h, a 50 µL aliquot was removed and processed in an identical manner to the analytical-scale reactions described above to determine substrate conversion. The remaining reaction mixture was diluted with acetone (2 x total reaction volume). Precipitated biomolecules were pelleted by centrifugation (4,000 x g, 12 min). Isolation procedure: The supernatant was concentrated under reduced pressure to a final volume of approximately 2 mL. The resulting mixture was filtered through a 0.22 µm filter and purified by preparative HPLC using a Phenomenex Kinetex 5 µm C18, 150 x 21.2 mm column under the following conditions: mobile phase A = deionized water + 0.1% formic acid and B = acetonitrile + 0.1% formic acid; method = 5% to 100% B over 13 min, 100% B for 4 min; flow rate, 15 mL/min.

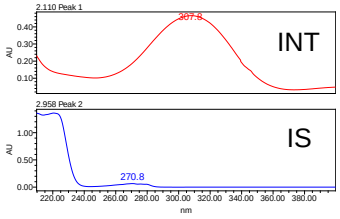
VII. UPLC Traces of Biotransformations

Figure 2.11. Oxidative dearomatization of **2.S1** by AfoD and AzaH. PDA traces of enzymatic reaction and control reaction.



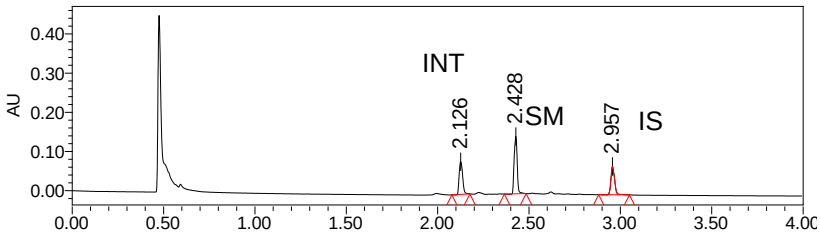
With AzaH

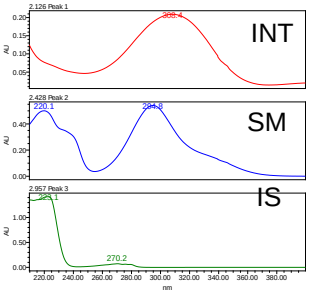




	Retention Time	Area	% Area	Height
1	2.110	305704	76.28	183131
2	2.957	95087	23.72	65154

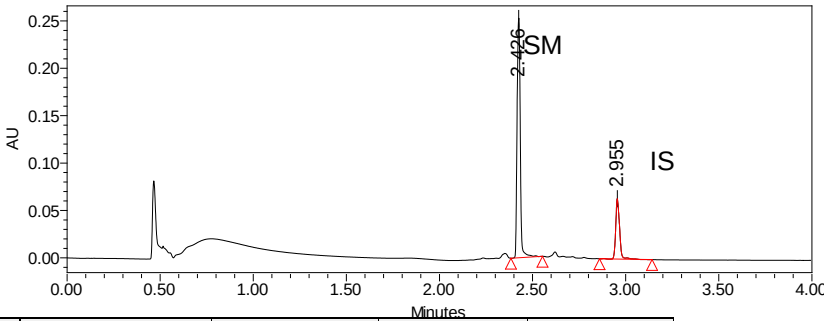
With AfoD

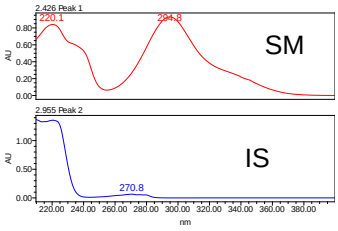




9	Retention Time	Area	% Area	Height
1	2.126	111917	29.11	83638
2	2.428	171664	44.65	146387
3	2.957	100880	26.24	72085

NEC



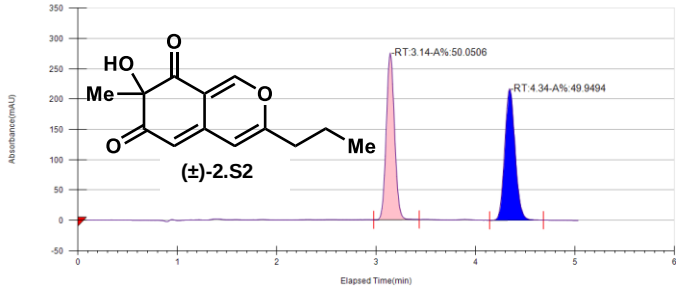


	Retention Time	Area	% Area	Height
1	2.426	301558	76.61	252996
2	2.955	92051	23.39	63755

VIII. Determination of Enantiomeric Excess

Figure 2.12. PDA traces of racemic **2.S2** obtained from an IBX-mediated oxidative dearomatization, (*S*)- **2.S2** obtained from AfoD-mediated oxidative dearomatization, (*R*)-**2.S2** obtained from AzaH-mediated oxidative dearomatization, and **2.S2** obtained from AfoD Y118F mediated oxidative dearomatization (CHIRALPAK® AD-H, 30%, CO_2, 3.5 mL/min).

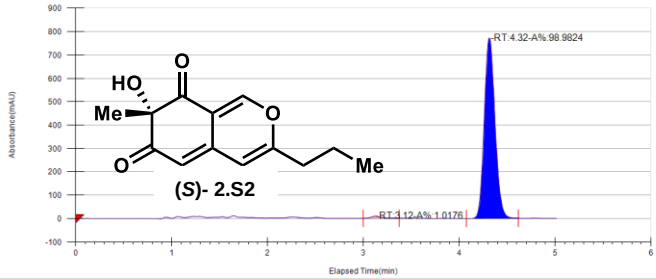
Racemic standard



Instrument method	Inj. Vol.	Solvent	Column	Sample	Temp.	Flow	% Modifier	Pressure
AD-H_30%_300-330	5	iPrOH	AD-H Chiral Analytical	sbdIV-087-rac	40	3.5	30	120

Peak No	% Area	Ret. Time	Cap Factor
1	50.0506	3.14 min	0
2	49.9494	4.34 min	0

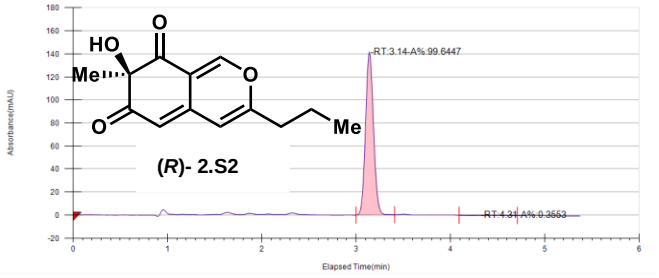
AfoD Reaction



Instrument method	Inj. Vol.	Solvent	Column	Sample	Temp.	Flow	% Modifier	Pressure
AD-H_30%_300-330	5	iPrOH	AD-H Chiral Analytical	sbdIV-087-AfoD	40	3.5	30	120

Peak No	% Area	Ret. Time	Cap Factor
1	1.0176	3.12 min	0
2	98.9824	4.32 min	0

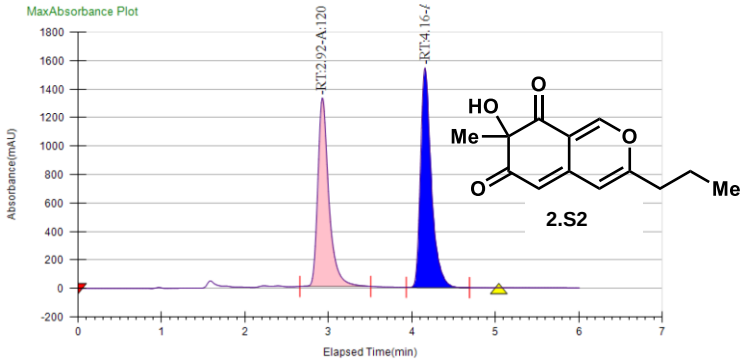
AzaH Reaction



Instrument method	Inj. Vol.	Solvent	Column	Sample	Temp.	Flow	% Modifier	Pressure
AD-H_30%_300-330	5	iPrOH	AD-H Chiral Analytical	sbdIV-087-rac	40	3.5	30	120

Peak No	% Area	Ret. Time	Cap Factor
1	99.6447	3.14 min	0
2	0.3553	4.31 min	0

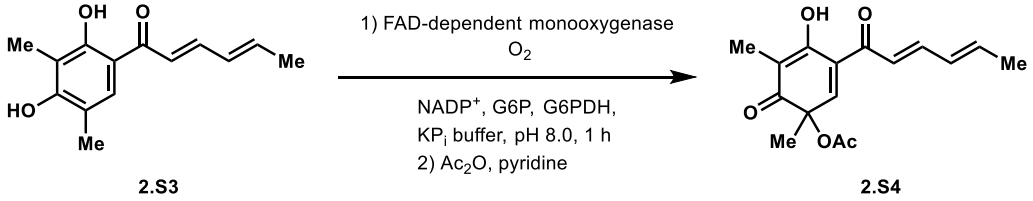
AfoD Y118F reaction



Instrument Method	Inj. Vol.	Solvent	Column	Sample	Well Location	Temp. (C)	Flow (g/min)	% Modifier	Pressure (Bar)
AD-H_30%_300-330	6 uL	Isopropanol	AD-H	ARB-V-071 AfoD_1	12A	40	3.5	30	120

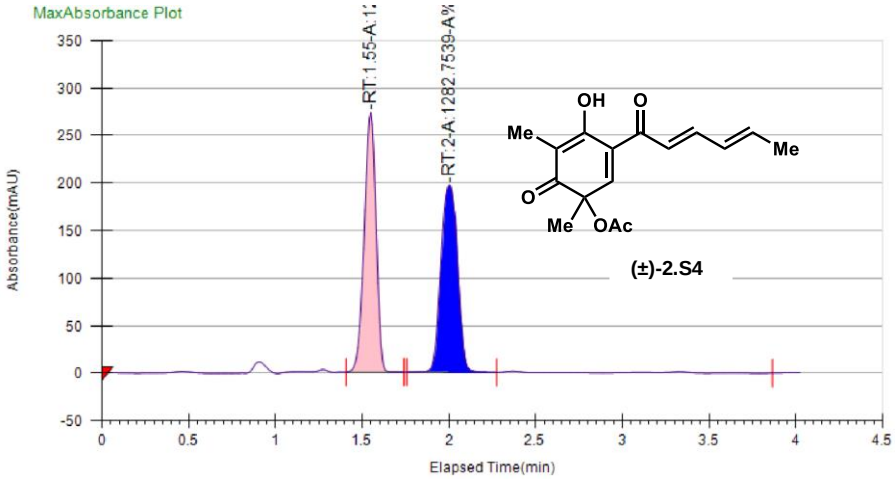
Peak No	% Area	Area	Ret. Time	Height	Cap. Factor
1	46.8333	12090.3845	2.92 min	1322.9644	0
2	53.1667	13725.3924	4.16 min	1541.5059	0

Figure 2.13. PDA traces of racemic **2.S4** obtained from an IBX-mediated oxidative dearomatization and (*S*)- **2.S4** obtained from FDMO7-mediated oxidative dearomatization (CHIRALPAK® AD-H, 30%, CO_2, 3.5 mL/min).



2.S3 **2.S4**

Racemic standard



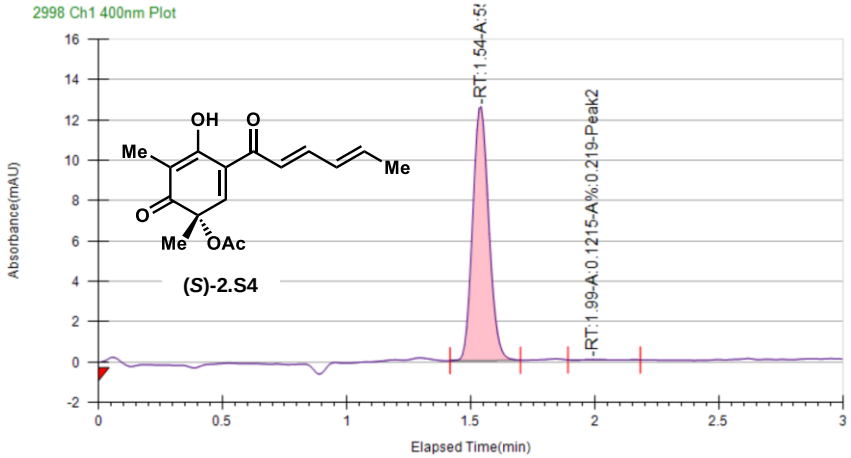
Run Information

Instrument Method	Inj. Vol.	Solvent	Column	Sample	Well Location	Temp.	Flow	% Modifier	Pressure
AD-H_10%	10	iPrOH	AD-H Chiral Analytical	sbdIV-003_ADH	15A	40	3.5	10	120

Peak Information

Peak No	% Area	Area	Ret. Time	Height	Cap. Factor
1	49.5405	1259.3926	1.55 min	273.3204	1548.9833
2	50.4595	1282.7539	2 min	197.3901	1998.9833

FDMO-7 reaction



Run Information

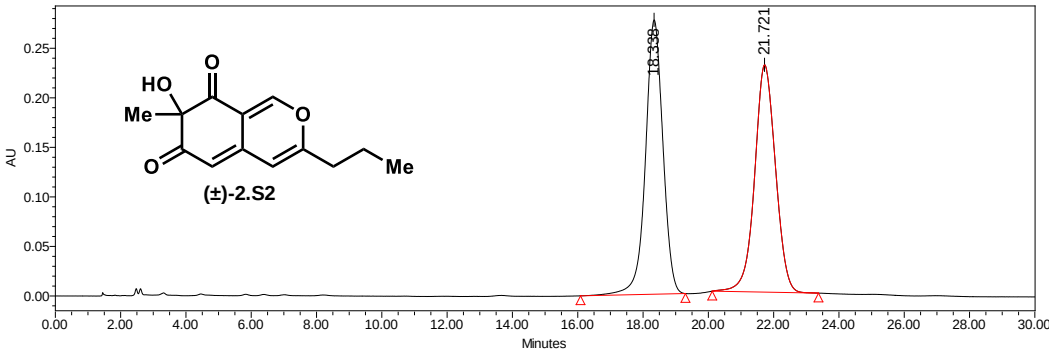
Instrument Method	Inj. Vol.	Solvent	Column	Sample	Well Location	Temp.	Flow	% Modifier	Pressure
AD-H_10%	1	iPrOH	AD-H Chiral Analytical	sbdVIII-120	11A	40	3.5	10	120

Peak Information

Peak No	% Area	Area	Ret. Time	Height	Cap. Factor
1	99.781	55.3487	1.54 min	12.5858	0
2	0.219	0.1215	1.99 min	0.0203	0

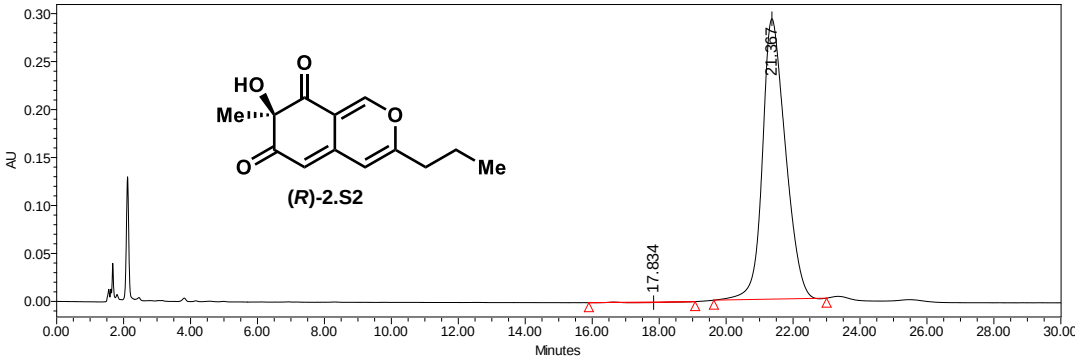
Figure 2.14. PDA traces of racemic **2.S2** obtained from an IBX-mediated oxidative dearomatization, (*R*)- **2.S2** obtained from FDMO-2 and FDMO-5-mediated oxidative dearomatization, and (*S*)- **2.S2** obtained from FDMO-6-mediated oxidative dearomatization (Phenomenex Lux Cellulose, 25% MeCN, 75% H_2O, 1 mL/min).

Racemic standard



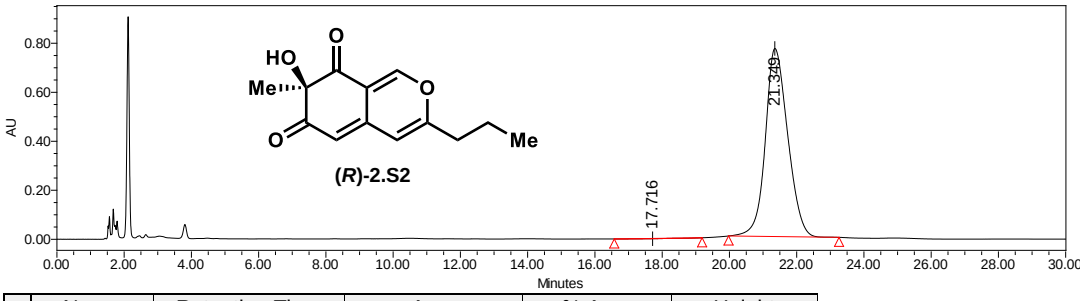
	Name	Retention Time	Area	% Area	Height
1		18.341	13429370	50.10	362068
2		21.724	13377480	49.90	300570

FDMO-2 reaction



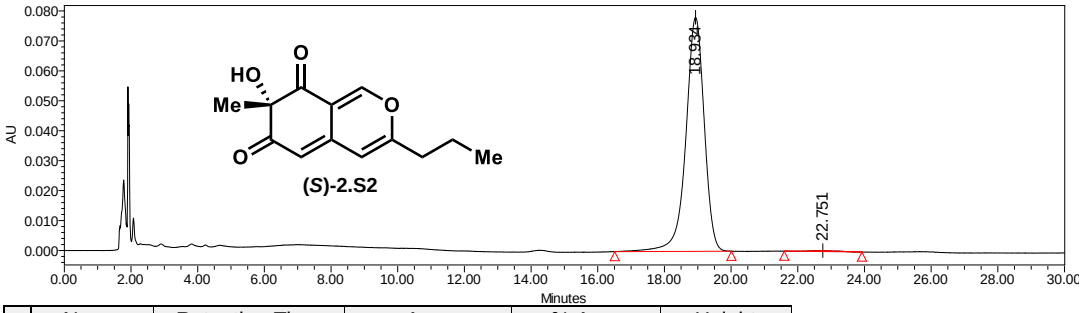
	Name	Retention Time	Area	% Area	Height
1		17.860	229964	0.45	-2129
2		21.365	51102440	99.55	1059976

FDMO-5 reaction



	Name	Retention Time	Area	% Area	Height
1		18.829	149446	0.42	-2017
2		21.349	35753101	99.58	769344

FDMO-6 reaction



	Name	Retention Time	Area	% Area	Height
1		18.934	3146288	99.83	83349
2		22.126	5282	0.17	-105

2.6 References

(1) Devine, P. N.; Howard, R. M.; Kumar, R.; Thompson, M. P.; Truppo, M. D.; Turner, N. J., Extending the application of biocatalysis to meet the challenges of drug development. *Nat. Rev. Chem.* **2018**, *2*, 409-421.

(2) France, S. P.; Aleku, G. A.; Sharma, M.; Mangas-Sanchez, J.; Howard, R. M.; Steflik, J.; Kumar, R.; Adams, R. W.; Slabu, I.; Crook, R.; Grogan, G.; Wallace, T. W.; Turner, N. J., Biocatalytic Routes to Enantiomerically Enriched Dibenz[c,e]azepines. *Angew. Chem. Int. Ed.* **2017**, *56*, 15589-15593.

(3) Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J., Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science* **2010**, *329*, 305-309.

(4) Bajaj, P.; Sreenilayam, G.; Tyagi, V.; Fasan, R., Gram-Scale Synthesis of Chiral Cyclopropane-Containing Drugs and Drug Precursors with Engineered Myoglobin Catalysts Featuring Complementary Stereoselectivity. *Angew. Chem. Int. Ed*. **2016**, *55*, 16110-16114.

(5) Loskot, S. A.; Romney, D. K.; Arnold, F. H.; Stoltz, B. M., Enantioselective Total Synthesis of Nigelladine A via Late-Stage C–H Oxidation Enabled by an Engineered P450 Enzyme. *J. Am. Chem. Soc.* **2017,** *139*, 10196-10199.

(6) Truppo, M. D., Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med. Chem. Lett.* **2017**, *8*, 476-480.

(7) Hernandez, K. E.; Renata, H.; Lewis, R. D.; Kan, S. B. J.; Zhang, C.; Forte, J.; Rozzell, D.; McIntosh, J. A.; Arnold, F. H., Highly Stereoselective Biocatalytic Synthesis of Key Cyclopropane Intermediate to Ticagrelor. *ACS Catal.* **2016**, *6*, 7810-7813.

(8) FDA'S policy statement for the development of new stereoisomeric drugs. *Chirality* **1992**, *4*, 338-340.

(9) Beletskaya, I. P.; Nájera, C.; Yus, M., Stereodivergent Catalysis. *Chem. Rev.* **2018,** *11*8, 5080-5200.

(10) Liu, C.; Lin, Z. W.; Zhou, Z. H.; Chen, H. B., Stereodivergent synthesis of all the four stereoisomers of antidepressant reboxetine. *Org. Biomol. Chem.* **2017**, *15*, 5395-5401.

(11) Weinstock, M. T.; Jacobsen, M. T.; Kay, M. S., Synthesis and folding of a mirror-image enzyme reveals ambidextrous chaperone activity. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 11679-84.

(12) Vinogradov, A. A.; Evans, E. D.; Pentelute, B. L., Total synthesis and biochemical characterization of mirror image barnase. *Chem. Sci.* **2015**, *6*, 2997-3002.

(13) Baker Dockrey, S. A.; Lukowski, A. L.; Becker, M. R.; Narayan, A. R. H., Biocatalytic site- and enantioselective oxidative dearomatization of phenols. *Nat. Chem.* **2018**, *10*, 119-125.

(14) Nicolaou, K. C.; Vassilikogiannakis, G.; Simonsen, K. B.; Baran, P. S.; Zhong, Y.-L.; Vidali, V. P.; Pitsinos, E. N.; Couladouros, E. A., Biomimetic Total Synthesis of Bisorbicillinol, Bisorbibutenolide, Trichodimerol, and Designed Analogues of the Bisorbicillinoids. *J. Am. Chem. Soc.* **2000**, *122*, 3071-3079.

(15) Bosset, C.; Coffinier, R.; Peixoto, P. A.; El Assal, M.; Miqueu, K.; Sotiropoulos, J.-M.; Pouységu, L.; Quideau, S., Asymmetric Hydroxylative Phenol Dearomatization Promoted by Chiral Binaphthylic and Biphenylic Iodanes. *Angew. Chem. Int. Ed.* **2014**, *53*, 9860-9864.
(16) Gerlt, J. A., Tools and strategies for discovering novel enzymes and metabolic pathways. *Perspect. Sci.* **2016**, *9*, 24-32.
(17) a) Stierle, A. A.; Stierle, D. B.; Bugni, T. Sequoiatones A and B: Novel Antitumor Metabolites Isolated from a Redwood Endophyte. *J. Org. Chem.* **1999**, *64*, 5479. b) Yoshida, E.; Fujimoto, H.: Baba, M.; Yamazaki, M. Four New Chlorinated Azaphilones, Helicusins A-D, Closely Related to 7-*epi*-Sclerotiorin, from an Ascomycetous Fungus, Talaromyces helicus. *Chem. Pharm. Bull.* **1995**, *43*, 1307. c) Quang, D. N.; Stadler, M.; Fournier, J.; Tomita, A.; Hashimoto, T. Cohaerins C–F, four azaphilones from the xylariaceous fungus Annulohypoxylon cohaerens. *Tetrahedron* **2006**, *62*, 6349. d) Laakso, J. A.; Raulli, R.; McElhaney-Feser, G. E.; Actor, P.; Underiner, T. L.; Hotovec, B. J.; Mocek, U.; Cihlar, R. L.; Broedel, S. E., Jr. CT2108A and B: New Fatty Acid Synthase Inhibitors as Antifungal Agents. *J. Nat. Prod.* **2003**, *66*, 1041 e) Nam, J. Y.; Son, K. H.; Kim, H. K.; Han, M. Y.; Kim, S. U.; Choi, J. D.; Kwon, B. M. Fungal pyranose oxidases: occurrence, properties and biotechnical applications in carbohydrate chemistry. *J. Microbiol. Biotechnol.* **2000**, *10*, 544.
(18) Kaur, K.; Wu, X.; Fields, J. K.; Johnson, D. K.; Lan, L.; Pratt, M.; Somoza, A. D.; Wang, C. C. C.; Karanicolas, J.; Oakley, B. R.; Xu, L.; De Guzman, R. N., The fungal natural product azaphilone-9 binds to HuR and inhibits HuR-RNA interaction in vitro. *PLOS ONE* **2017**, *12*, e0175471.
(19) Wang, W.; Liao, Y.; Chen, R.; Hou, Y.; Ke, W.; Zhang, B.; Gao, M.; Shao, Z.; Chen, J.; Li, F., Chlorinated Azaphilone Pigments with Antimicrobial and Cytotoxic Activities Isolated from the Deep Sea Derived Fungus Chaetomium sp. NA-S01-R1. *Mar. Drugs* **2018**, *16*.
(20) Yasukawa, K.; Takahashi, M.; Natori, S.; Kawai, K. i.; Yamazaki, M.; Takeuchi, M.; Takido, M., Azaphilones Inhibit Tumor Promotion by 12-O-Tetradecanoylphorbol-13-Acetate in Two-Stage Carcinogenesis in Mice. *Oncology* **1994**, *51*, 108-112.
(21) Matsuzaki K, T. H., Inokoshi J, Tanaka H, Masuma R,; S., O., New brominated and halogen-less derivatives and structure-activity relationship of azaphilones inhibiting gp120-CD4 binding. *J. Antibiot.* **1998**, *51*, 1004-1011.
(22) Tang, J.-L.; Zhou, Z.-Y.; Yang, T.; Yao, C.; Wu, L.-W.; Li, G.-Y., Azaphilone Alkaloids with Anti-inflammatory Activity from Fungus Penicillium sclerotiorum cib-411. *J. Agric. Food Chem.* **2019**, *67*, 2175-2182.
(23) Hsu, W.-H.; Chen, T.-H.; Lee, B.-H.; Hsu, Y.-W.; Pan, T.-M. Monascin and ankaflavin act as natural AMPK activators with PPARα agonist activity to down-regulate nonalcoholic steatohepatitis in high-fat diet-fed C57BL/6 mice. *Food Chem. Toxicol.* **2014**, *64*, 94-103.
(24) Udagawa, S., (-)-Sclerotiorin, A Major Metabolite of Penicillium hirayamae. *Chem. Pharm. Bull.* **1963,** *11*, 366-367.

(25) Whalley, W. B.; Ferguson, G.; Marsh, W. C.; Restivo, R. J., The chemistry of fungi. Part LXVIII. The absolute configuration of (+)-sclerotiorin and of the azaphilones. *J. Chem. Soc., Perkin Trans. 1* **1976**, *13*, 1366-1369.
(26) Nukina, M.; Marumo, S., Lunatoic acid A and B, aversion factor and its related metabolite of cochliobolus lunata. *Tetrahedron Lett.* **1977**, *18*, 2603-2606.
(27) Thines, E.; Anke, H.; Sterner, O., Trichoflectin, a Bioactive Azaphilone from the Ascomycete Trichopezizella nidulus. *J. Nat. Prod.* **1998**, *61*, 306-308.
(28) Park, J.-H.; Choi, G. J.; Jang, K. S.; Lim, H. K.; Kim, H. T.; Cho, K. Y.; Kim, J.-C., Antifungal activity against plant pathogenic fungi of chaetoviridins isolated from Chaetomium globosum. *FEMS Microbiol. Lett.* **2005,** *252*, 309-313.
(29) Germain, A. R.; Bruggemeyer, D. M.; Zhu, J.; Genet, C.; O'Brien, P.; Porco, J. A., Synthesis of the azaphilones (+)-sclerotiorin and (+)-8-O-methylsclerotiorinamine utilizing (+)-sparteine surrogates in copper-mediated oxidative dearomatization. *J. Org. Chem.* **2011**, *76*, 2577-2584.
(30) Zallot, R.; Oberg, N. O.; Gerlt, J. A., 'Democratized' genomic enzymology web tools for functional assignment. *Curr. Opin. Chem. Biol.* **2018**, *47*, 77-85.
(31) Gerlt, J. A., Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence–Function Space and Genome Context to Discover Novel Functions. *Biochemistry* **2017,** *56*, 4293-4308.
(32) Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L., Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta. Proteins Proteom.* **2015**, *1854*, 1019-1037.
(33) Chiang, Y.-M.; Szewczyk, E.; Davidson, A. D.; Keller, N.; Oakley, B. R.; Wang, C. C. C., A Gene Cluster Containing Two Fungal Polyketide Synthases Encodes the Biosynthetic Pathway for a Polyketide, Asperfuranone, in Aspergillus nidulans. *J. Am. Chem. Soc.* **2009**, *131*, 2965-2970.
(34) Davison, J.; al Fahad, A.; Cai, M.; Song, Z.; Yehia, S. Y.; Lazarus, C. M.; Bailey, A. M.; Simpson, T. J.; Cox, R. J., Genetic, molecular, and biochemical basis of fungal tropolone biosynthesis. *Proc. Natl. Acad. Sci.* **2012**, *109*, 7642-7647.
(35) Somoza, A. D.; Lee, K.-H.; Chiang, Y.-M.; Oakley, B. R.; Wang, C. C. C., Reengineering an azaphilone biosynthesis pathway in Aspergillus nidulans to create lipoxygenase inhibitors. *Org. Lett.* **2012**, *14*, 972-975.
(36) Rodríguez Benítez, A.; Tweedy, S. E.; Baker Dockrey, S. A.; Lukowski, A. L.; Wymore, T.; Khare, D.; Brooks, C. L.; Palfey, B. A.; Smith, J. L.; Narayan, A. R. H., Structural Basis for Selectivity in Flavin-Dependent Monooxygenase-Catalyzed Oxidative Dearomatization. *ACS Catal.* **2019**, *9*, 3633-3640.
(37) Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E., The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845.
(38) Dockrey, S. A. B.; Suh, C. E.; Benítez, A. R.; Wymore, T.; Brooks, C. L.; Narayan, A. R. H., Positioning-Group-Enabled Biocatalytic Oxidative Dearomatization. *ACS Cent. Sci.* **2019**, *5*, 1010-1016.

Chapter 3: Stereodivergent, Chemoenzymatic Synthesis of Azaphilone Natural Products

Summary:

Azaphilones represent a class of fungal natural products containing a highly oxygenated pyranoquinone bicyclic core and a tetrasubstituted carbon. While they often function as fungal pigments, many possess potent biological activity including anti-cancer, anti-microbial, and anti-inflammatory properties. Despite continued efforts towards their investigation, the biological activity of these compounds remains underexplored due to the synthetic challenges presented by their densely functionalized core and congested stereocenter. Having identified a number of biocatalysts that operate with complementary site- and stereoselectivity to access the azaphilone bicyclic core, the biocatalysts AzaH and AfoD were employed by Dr. Summer Bakery Dockrey and I towards the stereodivergent, chemoenzymatic synthesis of several azaphilone natural products. Specifically, the first syntheses of trichoflectin, deflectin-1a, and lunatoic acid A were

achieved. In addition, chemoenzymatic syntheses of these azaphilones supplied enantioenriched material for reassignment of the absolute configuration of trichoflectin and deflectin-1a based on optical rotation, CD spectra, and X-ray crystallography. These chemoenzymatic routes have simplified the stereoselective synthesis of azaphilones, laying the groundwork for constructing more complex azaphilones through multi-enzyme sequences.

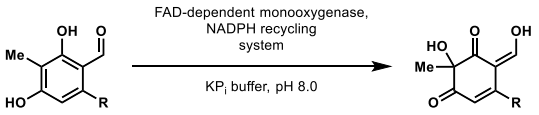
3.1 Introduction

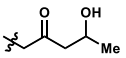
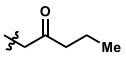
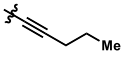
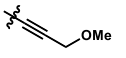
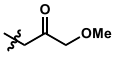
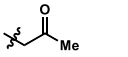
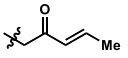
The ability to rapidly construct complex molecules with ease is a major hurtle in the identification and investigation of new biologically active compounds. Furthermore, with the increase in cases of antibiotic resistant infections and chemoresistant cancers, the need for highly selective and facile synthetic transformations is ever-growing.[1] To diversify these compounds for identification of drug leads, so called “modular synthesis” presents a unique approach to accessing the necessary breathe of chemical space. This method seeks to break down complex chemical structures into basic, interchangeable building blocks or modules, allowing for straightforward diversification strategies using highly-selective catalysts. This modularity enables chemists to more easily construct libraries of related compounds, providing a means to pinpoint which changes in the peripheral chemical structure increase activity towards a particular biological target through structure-activity relationship (SAR) studies.[1] However, although this approach promises insight into the details of minute changes to natural product scaffolds, the speed and precision with which current synthetic methods can be used to construct analogs is a significant bottleneck.

Benefitting from eons of evolution, the molecular machines responsible for generating natural product scaffolds are a logical starting point to begin the construction of more potent and selective drug candidates to combat current health crises. To overcome accessibility challenges in the context of azaphilone natural product synthesis, we sought to improve synthetic methodology towards these molecules using biocataylsis. Existing approaches for the construction of these molecules rely either on the cyclization of linear precursors[2] or, more commonly, the oxidative dearomatization of highly substituted resorcinol intermediates.[3-5] This biomimetic, dearomative strategy has proven valuable as it allows for the installation of structural complexity through well-established arene functionalization methods prior to the generation of the bicyclic core. Established approaches for oxidative dearomatization of azaphilone precursors have employed oxidants such as hypervalent iodide reagents[5-7] or $Pb(OAc)_4$.[3] Racemic syntheses of azaphilones have been reported using these reagents; however, achieving high levels of enantioselectivity can be a challenge. Enantioenriched azaphilones have been accessed using Porco and coworkers' copper-oxo (–)-sparteine dearomatization strategy,[8] achieving the total synthesis of (-)-mitorubrin with 97% ee,[9] as well as that of (+)-sclerotiorin and (+)-8-*O*-methylsclerotiorinamine from a common precursor obtained in 98% ee.[10] This method requires superstoichiometric quantities of both the oxidant and chiral ligand. To date, there has yet to be a report of a highly enantioselective, catalytic oxidative dearomatization using a small- molecule catalyst. Furthermore, the application of asymmetric oxidative dearomatization to the synthesis of tricyclic members of the azaphilone family, such as rubropunctatin and rubrotiorin, has yet to be reported. As such,

the absolute configuration of many of these natural products, which are typically assigned by analogy, has yet to be confirmed.[6, 11]

Table 3.1. Substrates screened for conversion by AfoD (0.8 mol %) and AzaH (0.8 mol %). *NADPH recycling system: G6P (2 equiv), NADP(+) (0.4 equiv), G6PDH (1 U/mL).



Entry	R	AfoD % conv.	AzaH % conv.
1		0%	73%
2		24%	64%
3		0%	48%
4		0%	>99%
5		19%	>99%
6		19%	94%
7		83%	95%

3.2 Identifying a Common Intermediate

Having demonstrated the ability to synthesize either enantiomer of azaphilone **2.S2** in Chapter 2, we sought to complete the stereodivergent syntheses of two biologically active azaphilone natural products. To identify a common intermediate that would provide versatility in accessing an array of azaphilone natural products, Dr. Baker Dockrey and I synthesized a number of aromatic substrates were subjected them to oxidative dearomatization with stereo-complementary biocatalysts. We hypothesized that the orcinaldehyde motif and homobenzylic ketone group would be critical for substrate recognition. In addition, we desired a functional handle to enable late-stage functionalization of the eastern portion of the azaphilone scaffold. Interestingly, the AzaH native substrate was not converted by AfoD (Table 3.1, entry 1). We hypothesize that the β-hydroxy group prevents the substrate from engaging in productive binding interactions and leads to the lack of activity of AfoD as the native substrate of AfoD features a hydrophobic group at this position.[12] This is supported by the activity of the saturated ketone substrate with AfoD (entry 2), which does not contain this polar functional group.

Promisingly, two alkynyl substrates that are intermediates in the synthesis of the ketone substrates are accepted by AzaH (entries 3 and 4); however, neither alkyne substrate was converted by AfoD. The α-methoxy ketone substrate (entry 5) was initially pursued as a promising intermediate as the dicarbonyl was fully converted by AzaH, but gave low conversions with AfoD. A simple methyl ketone derivative (entry 6) demonstrated activity with both enzymes as well; however, its low activity with AfoD ultimately disqualified it as the common substrate. The enone substrate depicted in entry 7 was efficiently dearomatized by both AfoD and AzaH in 83% and 95% conversion, respectively. The high conversions observed for this compound with both enzymes and the synthetic handle afforded by the double bond made it an ideal candidate for the common intermediate in our envisioned divergent synthesis of azaphilone natural products.

Despite its suitability in this role, the established route to this compound at the time involved a lengthy and often challenging ten step synthetic sequence.[8, 9] This route also relied on an expensive and volatile cross-coupling partner to install the alkene functionality at a late stage, which was undesirable. Therefore, to provide access to sufficient quantities of this enone substrate for further investigation, a new route was devised. Beginning from the commercially available and inexpensive reagent, methyl atratate, the phenolic groups on this compound were first MOM protected to yield **3.S4**. Weinreb amide **3.S5** was then synthesized and appended to **3.S4** to provide Weinreb ketone **3.S6** bearing the desired enone moiety. In spite of attempts to optimize this transformation, we achieved a maximum isolated yield of 30% of the desired product. We attribute this low yield to starting material consumption by the competing 1,4-addition

reaction of **3.S4** with the Weinreb amide, as opposed to the desired 1,2-addition. However, given that this step occurs early in our synthesis and can be performed on large scale, we deemed this yield sufficient for our purposes.

Having installed the enone functionality, we then set about reducing the ester present in the starting compound to the desired aldehyde. First, the newly installed carbonyl was protected from reduction through a base-mediated cyclization, generating lactone **3.S7**. This lactone was then reduced to the corresponding lactol **3.S8**. Next, we envisioned a global MOM deprotection under acidic conditions, anticipating that lactol **3.S8** would also open to reveal the desired aldehyde and ketone of the final compound. No productive reaction was observed under conventional deprotection conditions, prompting us to investigate Lewis acids to mediate this transformation. We found that $NbCl_5$ was able to cleanly deprotect both phenols while simultaneously opening the lactol, furnishing the desired enone **3.4**. Rigorous purification of this material was required, as any remaining Lewis acid appeared to render the material inactive in the enzymatic transformation. Thus, a maximum isolated yield of 50% was achieved for this step.

Ultimately, this synthesis provided streamlined access to **3.4**, reducing the longest linear step count from a commercially available reagent to five steps from the previous ten (see Experimental for details on this synthesis). This sequence was subsequently used to provide material for our syntheses of azaphilone natural products.

3.3 Synthesis of Trichoflectin

Angular azaphilone, trichoflectin (**3.1**), was the first natural product target chosen as a model for route development that would enable the synthesis of more complex

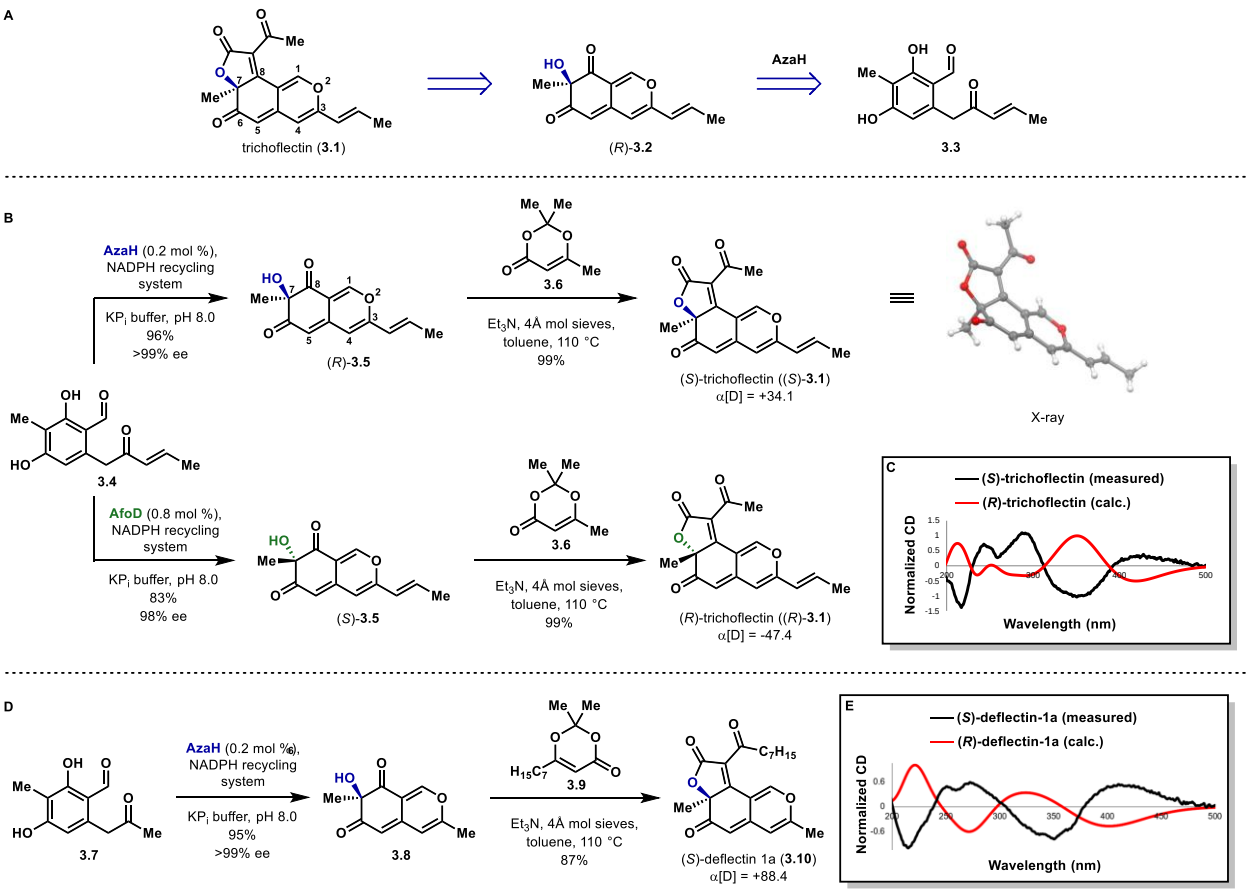


Figure 3.1. (A) Retrosynthetic analysis of the natural product trichoflectin (**1**). (B) Total synthesis of (*S*) and (*R*)-trichoflectin. (C) Calculated and measured CD data for (*S*) and (*R*)-trichoflectin. (D) Total synthesis of (*S*)-deflectin-1a (**10**). (E) Calculated and measured CD data for (*S*) and (*R*)-deflectin-1a. *NADPH recycling system: G6P (2 equiv), NADP(+) (0.4 equiv), G6PDH (1 U/mL).

tricyclic azaphilones. First isolated by Sterner in 1998, trichoflectin exhibits moderate antimicrobial activity as well as the ability to inhibit DHN-melanin biosynthesis in certain fungal species.[13] Despite interest in the biological activity of this compound, no total synthesis of trichoflectin has been reported to date. Our proposed retrosynthesis is described in Figure 3A. We envisioned constructing the butenolide ring of trichoflectin through an intramolecular Knoevenagel condensation following acylation of bicycle **3.2**.[6, 11] To obtain the reported (-)-trichoflectin, the C7-stereocenter would be set by stereoselective oxidative dearomatization of enone **3.4** with AzaH.

The synthesis of trichoflectin was initiated through a five-step route to enone **3.4** (see Experimental). Dearomatization of **3.4** with 0.2 mol % AzaH afforded azaphilone

scaffold (*R*)- **3.5** in 96% yield and >99% ee (Figure 3.3B). Next, acylation of (*R*)- **3.5** with the acylketene generated *in situ* from precursor **3.6** followed by Knoevenagel condensation provided (*S*)-trichoflectin ((*S*)- **3.1**) in 99% yield. This level of selectivity was not expected as Franck and coworkers reported the isolation of both angular and linear tricycles when constructing the butenolide ring of a similar azaphilone scaffold, proposing the distribution of these products is controlled by the steric properties of the acyl group.[11] However, in our system, the desired angular tricycle was produced exclusively. We attribute this to a difference in electronic properties of the extended π-system and have observed through attempts to reduce the C8-ketone that this position is more electrophilic than the C6-ketone.

Unexpectedly, the sign of the optical rotation of our synthetic (*S*)- **3.1** was opposite to that which is reported for the natural product (measured +34.1 $CHCl_3$, reported -121 $CHCl_3$).[13] To rule out the possibility of stereocenter inversion during the construction of the butenolide ring or that the enzyme performed the dearomatization on this particular substrate with unanticipated facial selectivity, a CD spectrum was calculated by Dr. Leo Joyce and compared to the measured spectrum of synthetic (*S*)- **3.1**. As depicted in Figure 3.1C, the calculated CD spectrum for (*R*)-trichoflectin is equal and opposite to the measured spectrum from (*S*)-trichoflectin, initially suggesting that the C7-stereocenter had been installed as anticipated. Ultimately, an X-ray crystal structure was obtained of the AzaH-produced tricycle by Dr. Ren Wiscons, unambiguously confirming the C7-configuration of (+)-trichoflectin as *S* (Figure 3.1B). Based on this data, we suspected that the natural product was misassigned. A 1976 report, which has been the basis for assignment of absolute configuration in the case of trichoflectin and many other

azaphilone natural products, suggests that the optical rotation of these compounds is controlled solely by the C7-stereocenter.[14] However, we have found that the electronic properties of the azaphilone core, as well as the presence of other stereocenters in the compound, can have a significant impact on the optical rotation based on computational modeling (see Experimental). To confirm the configuration of the natural product as *R* rather than *S* as assigned upon isolation, the enantiomeric tricycle was synthesized from the AfoD generated product (*S*)-**3.5** (Figure 3.1B). The optical rotation of (*R*)- **3.1** was measured as -47.4, agreeing in sign with the characterization of the isolated natural product. Therefore, it is proposed that the structure of the natural product be revised to the *R*-configuration.

3.4 Synthesis of Deflectin-1a

This revision prompted a careful inspection of the absolute configuration of structurally related angular tricycles. An azaphilone series called the deflectins, which are reported to possess inhibitory activity of bacteria and erythrocytes as well as cytotoxicity towards carcinoma cells, share the same tricyclic core as trichoflectin (Figure 3.1D).[15] Upon isolation, the absolute C7-configuration was assigned solely by optical rotation, as was done in the case of **3.1**.[13] Comparison of the calculated CD spectrum produced by Dr. Joyce of deflectin-1a to that of trichoflectin showed a good correlation (see Experimental, Figures 3.13 and 3.19), indicating that these compounds should exhibit similar optical properties. The deflectins are reported to possess the *R*-configuration at C7 and a corresponding optical rotation in the positive direction. This assignment

disagrees with the data collected for the trichoflectin enantiomers. Thus, it was deemed likely that this series was also assigned incorrectly.

To answer the question surrounding the absolute configuration of these molecules, we sought to complete the total synthesis of (*S*)-deflectin-1a (**3.10**, Figure 3.1D). As this compound contains a methyl group in place of the 1,2-disubstituted double bond found in trichoflectin, methyl ketone **3.7** was dearomatized with AzaH to produce bicycle **3.8** in 95% yield and >99% ee. Although AfoD also accepts **3.7** as a substrate (Table 3.1, entry 6), AzaH was used due to higher turnover and isolated yield in this reaction. Acylation and subsequent Knoevenagel condensation with the acylketene derived from precursor **3.9** furnished the desired butenolide to deliver (*S*)-deflectin-1a in 87% yield. The spectral data for **3.10** matched all available reported values. A measured optical rotation of +88.4 for this *S*-enantiomer confirms the need for structural revision of the natural product from the *R*-configuration to the *S* at C7.

3.5 Syntheisis of Lunatoic Acid A

With both trichoflectin and deflectin-1a in hand, we set out to construct the natural product lunatoic acid A (**3.11**), which was reported to possess the opposite configuration at C7 to the newly established configuration of natural trichoflectin and deflectin-1a. Isolated by Marumo, this natural product exhibited interesting antibacterial properties, as well as the ability to act as an antifungal agent by inducing chlamydospore-like cells in certain fungi.[16, 17] As depicted by the retrosynthesis in Figure 3.2A, we envisioned accessing the carboxylic acid moiety of **3.11** through a cross metathesis with the C9,10 double bond of intermediate **3.12**.[9] We anticipated that the chiral aliphatic ester of **3.12**

could be constructed through an acylation of the C7-hydroxyl group of (*R*)- **3.5**, which would be accessed through dearomatization of common intermediate **3.4** with AzaH.

Having established a robust method for biocatalytic dearomatization of enone **3.4**, the first challenge we encountered en route to lunatoic acid A was the acylation of tertiary alcohol **5**. While the acylation had proceeded smoothly with the ketene en route to trichoflectin and deflectin-1a, our attempts at acylation of **3.5** using acyl chlorides or symmetric anhydrides consistently failed to produce chiral ester **3.12**, affording only starting material or under more forcing conditions, complete decomposition of **3.5**. Ultimately, we discovered that Yamaguchi esterification with **3.13** (Figure 3.2B) did afford the desired product, although after optimization of reaction conditions, a maximum 50% yield was obtained.[18] Mixed anhydride **3.13** was generated *in situ* from the corresponding

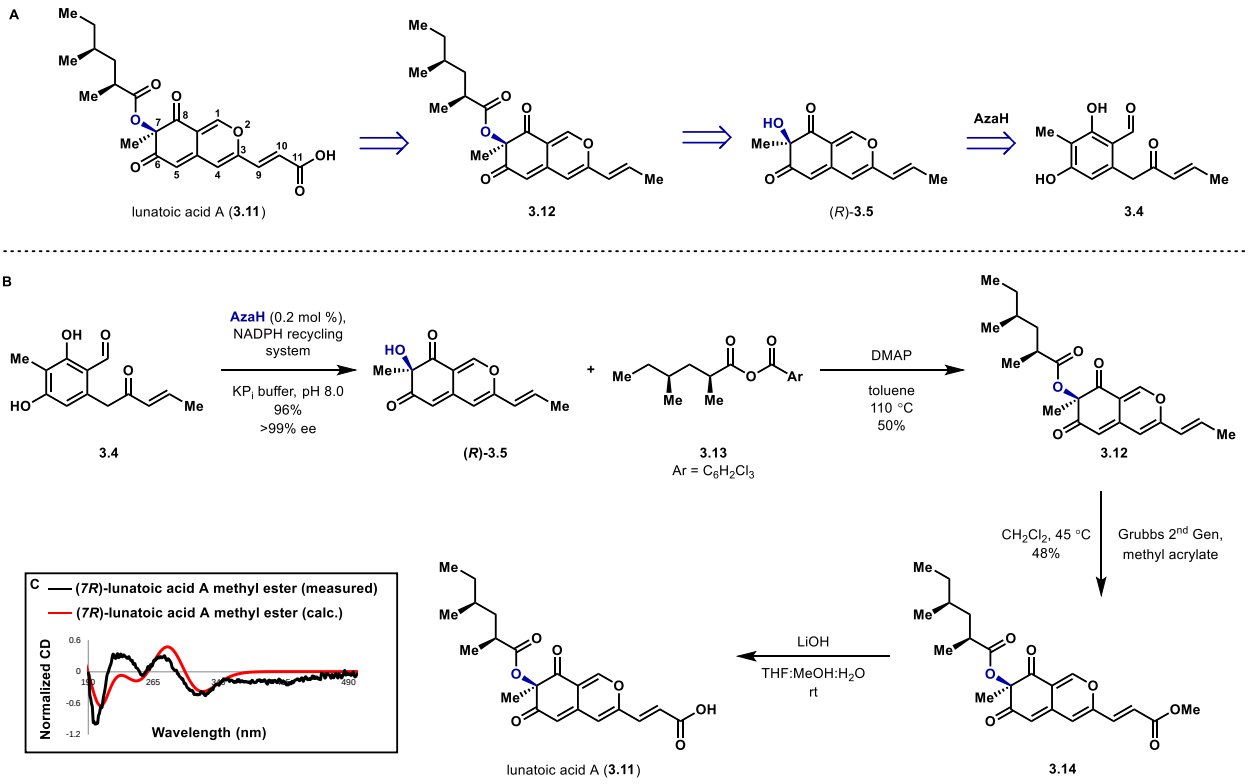


Figure 3.2 A) Retrosynthetic analysis of azaphilone natural product lunatoic acid A (**3.11**). B) Total synthesis of lunatoic acid A (**3.11**). C) Calculated and measured CD data for lunatoic acid A methyl ester* (**3.14**). *NADPH recycling system: G6P (2 equiv), NADP(+) (0.4 equiv), G6PDH (1 U/mL)

carboxylic acid, prepared as previously reported.[19] The carboxylic acid moiety present in lunatoic acid A was appended as the masked methyl ester through olefin metathesis with methyl acrylate and Grubbs 2nd generation catalyst in 48% yield.[20] Our attempts to perform the metathesis prior to esterification resulted in decomposition of the starting material, indicating that protection of the C7-hydroxyl group increases the stability of the compound to the reaction conditions. The expected mass of lunatoic acid A was observed by TOF-MS upon saponification of methyl ester **3.14** using LiOH; however, sufficient quantities of the pure compound for NMR studies could not be obtained. This observed instability parallels the nature of the free acid from culture extracts. These extracts were methylated to obtain lunatoic acid A methyl ester (**3.14**), which was fully characterized in lieu of the free acid.[16] All spectral data from the fully synthetic **3.14** match with reported literature values, including an agreement in the sign of the optical rotation and Cotton effects from the reported CD, indicating in this case that the original assignment of *R* at C7 is accurate.[16]

3.6 Conclusion

A set of complementary biocatalysts for oxidative dearomatization have been identified and applied to the enantiodivergent, chemoenzymatic synthesis of azaphilone natural products. A focused substrate library was created to identify a suitable common intermediate for dearomatization by both AzaH and AfoD. These enzymes were then used to construct both enantiomers of scaffold **3.5** to access the desired natural products containing either the *R*-or *S*-configuration at C7. The AzaH-produced scaffold (*R*)- **3.5** was initially used to access the natural product trichoflectin ((*S*)- **3.1**); however, it was

discovered that this tricycle contained the opposite stereocenter from the natural product based on its optical rotation, CD spectral data, and X-ray crystal structure. Synthesis of the opposite enantiomer of trichoflectin ((*R*)- **3.1**) was achieved using the dearomatized scaffold accessed through AfoD, with an optical rotation that led to the reassignment of the C7-stereocenter of trichoflectin to *S* from the original literature report of *R*. This prompted the construction of related tricycle deflectin-1a (**3.10**) using analogous substrates, which allowed for the reassignment of its C7-stereocenter. Lunatoic acid A (**3.11**) was then constructed from (*R*)- **3.5** and its structure confirmed through CD spectra and optical rotation. We anticipate these methods will enable the rapid construction of libraries of natural and unnatural azaphilones.

3.7 Experimental

I. Substrate Synthesis

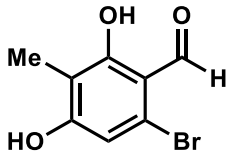
All reagents were used as received unless otherwise noted. Reactions were carried out under a nitrogen atmosphere using standard Schlenck techniques unless otherwise noted. Solvents were degassed and dried over aluminum columns on an MBraun solvent system (Innovative Technology, inc., Model PS-00-3). Reactions were monitored by thin layer chromatography using Millipore 60 F_{254} precoated silica TLC plates (0.25 mm) that were visualized using UV, *p*-anisaldehyde, CAM, DNP, or bromocresol stain. Flash column chromatography was performed using Machery-Nagel 60 μm (230-400 mesh) silica gel. All compounds purified by column chromatography were sufficiently pure for use in further experiments unless otherwise indicated. ^{1}H and ^{13}C NMR spectra were obtained in $CDCl_3$ at rt (25 °C), unless otherwise noted, on Varian 400 MHz or Varian 600

MHz spectrometers. Chemical shifts of ^{1}H NMR spectra were recorded in parts per million (ppm) on the δ scale. High resolution electrospray mass spectra were obtained on an Agilent HPLC-TOF at the University of Michigan Life Sciences Institute.

List of reagents prepared or purified:

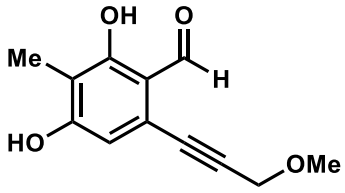
2-Iodoxybenzoic Acid (IBX) was synthesized according to the procedure described by Sputore *et al*.[21]

Trifluoroacetic acid and **acetic anhydride** were distilled prior to use.



6-Bromo-2,4-dihydroxy-3-methylbenzaldehyde (3.S1)

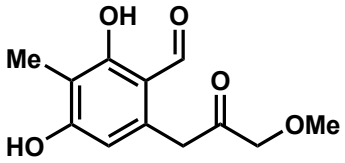
Prepared as previously reported by Baker Dockrey *et al*.[22]



2,4-Dihydroxy-6-(3-methoxyprop-1-yn-1-yl)-3-methylbenzaldehyde (3.S2)

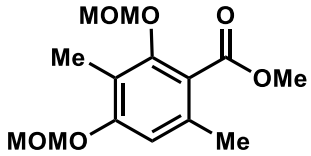
Aryl bromide **S1** (150 mg, 0.65 mmol), $PdCl_2(PPh_3)_2$ (23 mg, 0.033 mmol, 0.050 equiv), CuI (12 mg, 0.065 mmol, 0.10 equiv) was stirred in 4.8 mL of anhydrous DMF in a flame-dried round bottom flask equipped with a stir bar. Et_3N (0.30 mL, 2.2 mmol, 3.3 equiv) was added and the mixture was sparged with N_2 for 15 min before methyl propargyl ether (0.11 mL, 1.3 mmol, 2.0 equiv) was added. The resulting mixture was heated to 60 °C for

14 h. The reaction mixture was cooled to rt, diluted with water (2.0 mL), and acidified with 1 M HCl (4.0 mL). The mixture was extracted with EtOAc (3 x 10 mL) and the combined organic layers were washed with water and brine, dried over anhydrous Na_2SO_4, filtered, and concentrated under reduced pressure to afford a dark brown solid. Purification on silica gel (10-20% EtOAc in hexanes) afforded 135 mg (94% yield) of **S2** as a tan solid. **^{1}H NMR** (400 MHz, CD_3OD) δ 10.04 (s, 1H), 6.47 (s, 1H), 4.33 (s, 2H), 3.41 (s, 3H), 1.98 (s, 3H). All spectra obtained were consistent with literature values.[5]



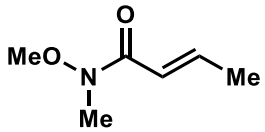
2,4-dihydroxy-6-(3-methoxy-2-oxopropyl)-3-methylbenzaldehyde (3.S3)

To a mixture of **3.S2** (20 mg, 0.091 mmol, 1.0 equiv) and $Au(OAc)_3$ (1.7 mg, 0.0046 mmol, 0.050 equiv) in DCE (1.0 mL) was added TFA (0.091 mL, 1.2 mmoL, 13 equiv) in a flame-dried vial under N_2. The resulting mixture was stirred for 1 h at rt before 10 mL of 20% H_2O/MeCN was added. The mixture was stirred for 12 h at rt before it was diluted with water (1.0 mL) and extracted with EtOAc (3 x 5.0 mL). The combined organic layers were washed with brine, dried over Na_2SO_4, and concentrated under reduced pressure to afford a dark brown solid. Purification on silica gel (80% EtOAc in hexanes) afforded 20 mg (92% yield) of **3.S3** as a light tan solid. **^{1}H NMR** (400 MHz, CD_3OD) δ 9.82 (s, 1H), 6.23 (s, 1H), f.02 (s, 2H), 3.39 (s, 3H), 2.00 (s, 3H);**^{13}C NMR** (150 MHz, CD_3OD) δ 205.8, 193.4, 163.7, 163.0, 137.0, 112.3, 110.2, 110.1, 76.4, 58.1, 41.0, 5.9; **HR-ESI-MS**: *m/z* calculated for $C_{12}H_{15}O_5$ $[M+H]^+$: 239.2465, found: 239.2465; **IR** (thin film): 2921, 2827, 1721, 1614, 1503 cm^{-1}; **MP**: 127-129 °C.



Methyl 2,4-bis(methoxymethoxy)-3,6-dimethylbenzoate (3.S4)

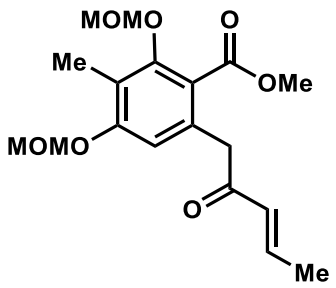
Methyl atratate (13 g, 66 mmol, 1.0 equiv) in THF (650 mL) was cooled to 0 °C and NaH (60%, 7.9 g, 200 mmol, 3.0 equiv) was added portionwise. MOMCl[23] (15 mL, 200 mmol, 3.0 equiv) was slowly added to the resulting mixture. The solution was warmed to rt and stirred for 5 h before it was cooled to 0 °C and quenched with NH_4Cl (500 mL, saturated aq.). The layers were separated, and the aqueous layer was extracted with EtOAc (3 x 500 mL). The combined organic layers were washed with $NaHCO_3$, dried over Na_2SO_4 and concentrated under reduced pressure to afford a yellow oil. Purification on silica gel (0-20% EtOAc in hexanes) afforded 15 g of the ester as a colorless oil (90% yield). **^{1}H NMR** (300 MHz, $CDCl_3$): δ 6.72 (s, 1H), 5.19 (s, 2H), 4.96 (s, 2H), 3.89 (s, 3H), 3.54 (s, 3H), 3.47 (s, 3H), 2.28 (s, 3H), 2.15 (s, 3H). All spectra obtained were consistent with literature values.[11]



(*E*)-N-methoxy-N-methylbut-2-enamide (3.S5)

Crotonic acid (10 g, 120 mmol, 1.0 equiv) was dissolved in oxalyl chloride (12 mL, 140 mmol, 1.2 equiv) and stirred at 70 °C for 1 h. The resulting acyl chloride was distilled at

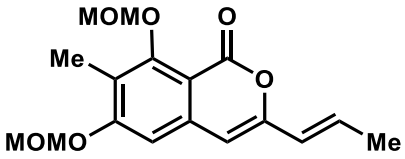
124 °C, then added to a solution of N,O-dimethylhydroxylamine hydrochloride (10 g, 100 mmol, 0.90 equiv) in DCM (190 mL). The mixture was cooled to 0 °C before pyridine (21 mL, 260 mmol, 2.2 equiv) was added slowly. The mixture was stirred at 0 °C for 30 min, then allowed to warm to rt for 30 min before it was diluted with 1 M HCl (150 mL) and extracted Et_2O (3 x 150 mL). The combined organic layers were washed with brine (300 mL), dried over $MgSO_4$, and concentrated to afford a red oil. Purification on silica gel (60-80% Et_2O in hexanes) afforded 12 g (82% yield) of the title compound as a pale yellow oil. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 6.98 (m, 1H), 6.41 (d, J = 15.4 Hz, 1H), 3.69 (s, 3H), 3.23 (s, 3H), 1.90 (d, J = 8.6 Hz, 3H). All spectra obtained were consistent with literature values.[24]



Methyl(*E*)-2,4-bis(methoxymethoxy)-3-methyl-6-(2-oxopent-3-en-1-yl)benzoate (3.S6)

Diisopropylamine (9.1 mL, 65 mmol, 1.1 equiv) in THF (300 mL) was stirred at -78 °C. *n*-BuLi (2.5 M in hexane, 26 mL, 65 mmol, 1.1 equiv) was slowly added. The resulting mixture was warmed to 0 °C and stirred for 15 min before it was cooled to -78 °C and a solution of ester **3.S4** (15 g, 20 mmol, 1.0 equiv) in THF (50 mL) was added. The resulting mixture was stirred at -78 °C for 15 min before a solution cooled to -78 °C of amide **3.S5** (9.2 g, 71 mmol, 1.2 equiv) in THF (25 mL) was added by cannula. The mixture was

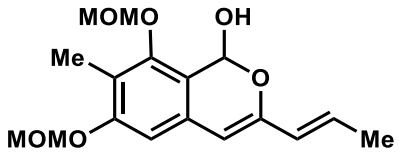
stirred for 1 h at -78 °C and was then acidified with 1 M HCl (aq. 100 mL). The layers were separated, and the aqueous layer was extracted with EtOAc (3 x 400 mL). The combined organic layers were washed with brine (400 mL), dried over Na_2SO_4, and concentrated under reduced pressure to afford a yellow oil. Purification on silica gel (5-15% EtOAc in hexanes) afforded 4.5 g of an inseparable mixture of the title compound and remaining amide **3.S5**. This mixture was carried forward without further purification.



(*E*)-6,8-Bis(methoxymethoxy)-7-methyl-3-(prop-1-en-1-yl)-1H-isochromen-1-one (3.S7)

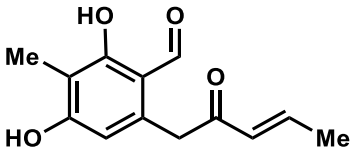
NaH (170 mg, 4.50 mmol, 1.05 equiv, 60%, dispersion in mineral oil) was stirred in THF (290 mL) at -20 °C. Enone **3.S6** (1.5 g, 4.3 mmol, 1.0 equiv) in THF (50 mL) was slowly added to the suspension. *t*-BuOH (10 µL) was added and the solution was stirred for 1 h at 20 °C. The reaction was cooled to 0 °C and quenched with EtOAc (30 mL), followed by addition of a saturated aqueous solution of NH_4Cl (40 mL). The mixture was then allowed to warm to rt. The layers were separated, and the aqueous layer extracted with EtOAc (3 x 200 mL). The combined organic layers were washed with brine (400 mL), dried over Na_2SO_4, and concentrated under reduced pressure to afford a white solid. Purification on silica gel (0-20% EtOAc in hexanes) afforded 1.3 g of the lactone **3.S7** as a white crystalline solid (22% yield over 2 steps). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 6.78 (s, 1H), 6.59 (m, 1H), 6.12 (s, 1H), 5.99 (dt, J = 15.7, 1.7 Hz, 1H), 5.28 (s, 2H), 5.16 (s, 2H), 3.64 (s, 3H), 3.49 (s, 3H), 2.27 (s, 3H), 1.89 (d, J = 6.9 Hz, 3H); **^{13}C NMR** (150 MHz,

$CDCl_3$) δ 161.0, 159.2, 158.9, 152.2, 139.1, 131.6, 122.9, 122.3, 107.4, 105.1, 103.7, 101.6, 94.2, 57.7, 56.4, 18.3, 9.9; **HR-ESI-MS**: *m/z* calculated for $C_{17}H_{21}O_6$ $[M+H]^+$: 321.1333, found: 321.1335; **IR** (thin film): 2934, 2832, 1715, 1658, 1598 cm^{-1}**; MP**: 124-126 °C.



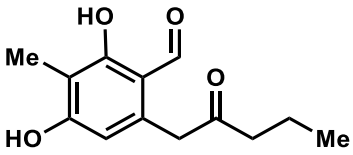
(*E*)-6,8-bis(methoxymethoxy)-7-methyl-3-(prop-1-en-1-yl)-1H-isochromen-1-ol (3.S8)

Lactone **3.S7** (1.0 g, 2.6 mmol, 1.0 equiv) in THF (13 mL) was stirred at -78 °C. DIBALH (0.48 mL, 2.7 mmol, 1.1 equiv) was added dropwise via syringe. The resulting solution was allowed to stir at -78 °C for 30 min, and then the reaction was quenched by the addition of EtOAc (1.0 mL) and Rochelle's salt (saturated, 4.0 mL). The resulting mixture was stirred for 1 h at rt. The crude mixture was extracted with EtOAc (3 x 50 mL), and the combined organic layers were washed brine (20 mL), dried over Na_2SO_4, and concentrated under reduced pressure. Purification on silica gel (0-20% EtOAc in hexanes) afforded 730 mg of the lactol **3.S8** as a colorless oil (89% yield). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 6.63 (s, 1H), 6.56 (d, J = 5.8 Hz, 1H), 6.33 (m, 1H), 5.97 (dd, J = 15.3, 2.1 Hz, 1H), 5.78 (s, 1H), 5.16 (s, 2H), 5.01 (s, 2H), 3.60 (s, 3H), 3.45 (s, 3H), 2.15 (s, 3H), 1.84 (m, 3H); **^{13}C NMR** (100 MHz, $CDCl_3$) δ 156.7, 153.0, 148.7, 129.1, 128.3, 126.0, 119.0, 116.4, 105.8, 102.0, 99.9, 94.4, 89.0, 57.5, 56.0, 18.2, 9.9; **HR-ESI-MS**: *m/z* calculated for $C_{17}H_{23}O_6$ $[M+H]^+$: 323.1489, found: 323.1489**; IR** (thin film): 3357, 2930, 2823, 1604, 1447 cm^{-1}; **MP**: 62-65 °C.



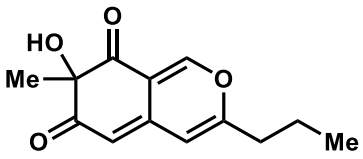
(*E*)-2,4-Dihydroxy-3-methyl-6-(2-oxopent-3-en-1-yl)benzaldehyde (3.4)

Lactol **3.S8** (300 mg, 0.933 mmol) in MeCN and water (7:1, 0.04 M) was stirred at rt. $LiBF_4$ (630 mg, 6.7 mmol, 7.2 equiv) was added and the resulting mixture was stirred at 70 °C for 3 h. The reaction mixture was cooled to rt and quenched by the addition of water (15 mL). The mixture was diluted with EtOAc (30 mL) and the layers were separated. The aqueous layer was extracted with EtOAc (3 x 30 mL) and the combined organic layers were washed brine (40 mL), dried over Na_2SO_4, and concentrated under reduced pressure. Purification on silica gel (10-40% EtOAc in hexanes) afforded 200 mg (90% yield) of the title compound as a tan crystalline solid. **^{1}H NMR** (600 MHz, CD_3OD) δ 9.77 (s, 1H), 7.06 (m, 1H), 6.25 (d, J = 1.8 Hz, 1H), 6.22 (s, 1H), 4.12 (s, 2H), 2.00 (s, 3H), 1.92 (d, J = 5.2 Hz, 3H); **^{13}C NMR** (150 MHz, CD_3OD) δ 197.5, 193.4, 163.7, 163.0, 144.8, 138.0, 130.3, 112.3, 110.3, 109.9, 42.5, 17.0, 5.8; **HR-ESI-MS**: *m/z* calculated for $C_{13}H_{15}O_4$ $[M+H]^+$: 235.0965, found: 235.0947**;** **IR** (thin film): 3231, 2928, 2824, 2409, 1614 cm^{-1}; **MP**: 139-142 °C.



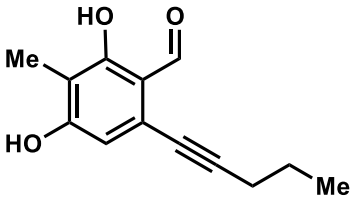
2,4-Dihydroxy-3-methyl-6-(2-oxopentyl)benzaldehyde (3.S9)

Prepared as previously reported by Baker Dockrey *et al*.[22]



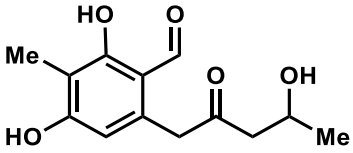
7-hydroxy-7-methyl-3-propyl-6H-isochromene-6,8(7H)-dione (3.S10)

Prepared as previously reported by Baker Dockrey *et al*.[2]



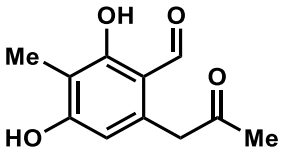
2,4-Dihydroxy-3-methyl-6-(pent-1-yn-1-yl)benzaldehyde (3.S11)

Prepared as previously reported by Baker Dockrey *et al*.[22]



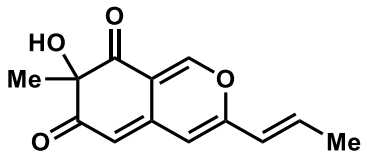
2,4-Dihydroxy-6-(4-hydroxy-2-oxopentyl)-3-methylbenzaldehyde (3.S12)

Prepared as previously reported by Baker Dockrey *et al*.[22]



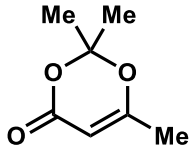
2,4-dihydroxy-3-methyl-6-(2-oxopropyl)benzaldehyde (3.7)

Prepared as previously reported by Baker Dockrey *et al*.[22]



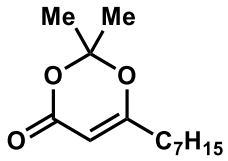
(*E*)-7-hydroxy-7-methyl-3-(prop-1-en-1-yl)-6*H*-isochromene-6,8(7*H*)-dione (3.v5)

26 (20 mg, 0.085 mmol, 1.0 equiv) and $Au(OAc)_3$ (1.7 mg, 0.0043 mmol, 0.05 equiv) in DCE (0.94 mL) were added to a flame-dried vial. The mixture was stirred at rt for 30 min before IBX (28 mg, 0.10 mmol, 1.1 equiv), TBAI (19 mg, 0.051 mmol, 0.60 equiv), and TFA (0.094 mL, 10% total volume) were added. The mixture was stirred for an additional 40 min before the reaction was quenched with 5 drops of a saturated $Na_2S_2O_3$ solution. The mixture was filtered through a plug of ceilte, which washed with DCM, before the solution was concentrated under reduced pressure. Purification by preparative TLC with 2.5% MeOH in DCM yielded 4.7 mg **5** (22% yield) of the title compound as an orange oil. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.90 (s, 1H), 6.59 (m, 1H), 6.10 (s, 1H), 6.01 (d, J = 15.6 Hz, 1H), 5.58 (s, 1H), 1.94 (d, J = 7.0 Hz, 3H), 1.55 (s, 3H). All spectra obtained were consistent with reported values.[9]



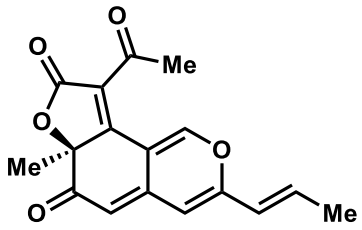
2,2,6-trimethyl-4H-1,3-dioxin-4-one (3.6)

Prepared as reported previously by Fuse *et al*.[25] 1.04 g (73% yield) of the title compound was obtained as a yellow oil. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 5.24 (s, 1H), 1.98 (s, 3H), 1.68 (s, 6H). All spectra obtained were consistent with literature values.[25]



6-heptyl-2,2-dimethyl-4H-1,3-dioxin-4-one (3.9)

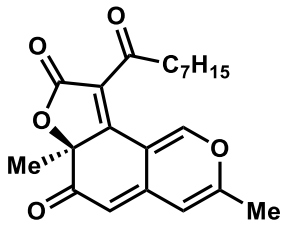
Prepared as reported previously by Franck and coworkers.[11] 356 mg (26% yield over 3 steps) of the title compound was obtained as a yellow oil and taken on crude without further purification.



trichoflectin (3.1)

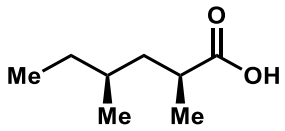
To a solution of **3.5** (5.5 mg, 0.023 mmol, 1.0 equiv) and dioxinone **3.6** (5 mg, 0.035 mmol, 1.5 equiv) in toluene (0.64 mL) in a flame dried vial under N_2 was added mol sieves. The mixture was stirred at rt for 10 min and then heated to 110 °C. After 1 h, Et_3N (0.0064 mL, 0.046 mmol, 2.0 equiv) was added. The mixture was stirred for an additional hour at 110 °C before it was cooled to room temperature and quenched with 1 M HCl (1.0 mL). The mixture was extracted with EtOAc (3 x 2.0 mL). The organic layers were combined, washed with brine, dried over Na_2SO_4, and concentrated under reduced pressure. Purification by preparative HPLC yielded 7.2 mg of **1** (>99% yield) as a yellow oil. **^{1}H NMR** (600 MHz, $CDCl_3$) δ 8.82 (s, 1H), 6.63 (m, 1H), 6.06 (s, 1H), 6.01 (d, J = 13.8 Hz, 1H), 5.35 (d, J = 1.2 Hz, 1H), 2.60 (s, 3H), 1.95 (d, J = 5.2 Hz, 3H), 1.69 (s, 3H); **^{13}C NMR** (150 MHz, $CDCl_3$) δ 194.5, 190.0, 168.2, 165.6, 155.2, 153.1, 144.0, 136.3, 123.3, 122.3,

110.9, 107.5, 105.8, 87.6, 30.1, 29.7, 26.3, 18.7; **HR-ESI-MS**: *m/z* calculated for $C_{17}H_{15}O_5$ $[M+H]^+$: 299.0914, found: 299.0922; **IR** (thin film): 2921, 2827, 1721, 1614, 1503 cm^{-1}; $[\alpha]_D$ +34 ° (*c* 0.1, $CHCl_3$). All spectra obtained were consistent with literature values.[13]



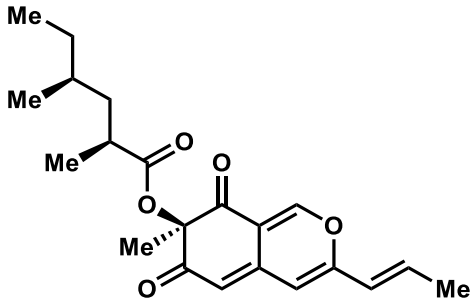
deflectin-1a (3.10)

To a solution of **3.8** (4.0 mg, 0.019 mmol, 1.0 equiv) and dioxinone **3.9** (6.6 mg, 0.029 mmol, 1.5 equiv) in toluene (0.55 mL) in a flame dried vial under N_2 was added mol sieves. The mixture was stirred at rt for 10 min and then was heated to 110 °C. After 1 h, Et_3N (53 µL, 0.038 mmol, 2.0 equiv) was added. The mixture was stirred for an additional hour at 110 °C before it was cooled to room temperature and quenched with 1 M HCl (1.0 mL). The mixture was extracted with EtOAc (3 x 2.0 mL) and the combined organic layers were washed with (5.0 mL), dried over Na_2SO_4, and concentrated under reduced pressure. Purification on silica gel (20% EtOAc in hexanes) yielded 5.9 mg of **3.10** (87% yield) as a yellow oil. **^{1}H NMR** (599 MHz, $CDCl_3$) δ 8.77 (s, 1H), 6.08 (s, 1H), 5.28 (s, 1H), 3.16 (m, 1H), 2.83 (m, 1H), 2.20 (s, 3H), 1.68 (s, 3H), 1.62 (m, 2H), 1.55 (s, 3H), 1.27 (m, 8H), 0.87 (m, 3H); **^{13}C NMR** (151 MHz, $CDCl_3$) δ 197.3, 190.3, 168.1, 165.2, 158.7, 153.3, 144.1, 123.6, 111.1, 108.4, 104.9, 87.6, 42.1, 31.6, 29.0, 29.0, 26.2, 23.4, 22.6, 19.4, 14.1; **HR-ESI-MS**: *m/z* calculated for $C_{21}H_{25}O_5$ $[M+H]^+$: 357.1697**,** found: 357.1754; **IR** (thin film): 2924, 1763, 1684, 1642 1540 cm^{-1}; $[\alpha]_D$ +88 ° (*c* 0.1, EtOAc). All spectra obtained were consistent with literature values.[15]



(2S,4S)-2,4-dimethylhexanoic acid (3.S13)

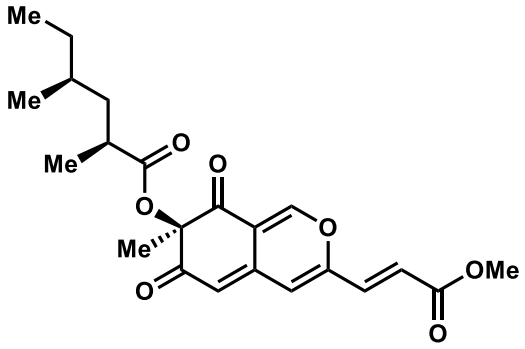
Prepared as reported previously by Myers *et al*.[19] 278 mg (32% yield over 3 steps) of the title compound was obtained as a white crystalline solid. **^{1}H NMR** (599 MHz, $CDCl_3$) δ 2.56 (m, 1H), 1.72 (m, 1H), 1.39 (m, 1H), 1.32 (m, 1H), 1.17 (d, J = 7.0 Hz, 3H), 1.13 (m, 2H), 0.88 (d, J = 6.6 Hz, 3H), 0.86 (t, J = 7.4 Hz, 3H). All spectra obtained were consistent with literature values.[19]



(R)-7-methyl-6,8-dioxo-3-((E)-prop-1-en-1-yl)-7,8-dihydro-6H-isochromen-7-yl (2S,4S)-2,4-dimethylhexanoate (3.12)

To a solution of carboxylic acid **3.S13** (4.5 mg, 0.031 mmol, 1.1 equiv) in THF (0.2 mL) was added 2,4,6-trichlorobenzoyl chloride (0.0048mL, 0.031 mmol, 1.1 equiv) and Et_3N (0.0043 mL, 0.031 mmol, 1.1 equiv). The mixture was stirred at rt for 30 min before it was filtered through celite and concentrated. The resulting clear oil was dissolved in toluene (0.8 mL) and added to **3.5** (6.5 mg, 0.028 mmol, 1 equiv) and DMAP (10.3 mg, 0.084 mmol, 3 equiv). This mixture was stirred at 110 °C for 1 h before it was cooled to rt, acidified with a half-saturated solution of NH_4Cl (1 mL), and extracted with EtOAc. The organic layers were washed with brine, dried over Na_2SO_4, and concentrated under

reduced pressure. Purification on silica gel (10-30% EtOAc in hexanes) provided 5 mg (50% yield) of the title compound as an orange oil. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.88 (s, 1H), 6.55 (m, 1H), 6.07 (s, 1H), 6.00 (d, J = 15.0 Hz, 1H), 5.57 (s, 1H), 2.70 (m, 1H), 1.93 (d, J = 6.9 Hz, 3H), 1.77 (m, 1H), 1.53 (s, 3H), 1.34 (m, 2H), 1.19 (d, J = 6.9 Hz, 3H), 1.13 (m, 2H), 0.89 (m, 6H). **^{13}C NMR** (151 MHz, $CDCl_3$) δ 193.2, 192.7, 176.3, 155.3, 153.4, 142.6, 135.3, 122.4, 114.9, 108.5, 107.8, 83.8, 40.8, 36.2, 31.8, 29.4, 22.1, 19.1, 18.6, 17.6, 11.1; **HR-ESI-MS**: *m/z* calculated for $C_{21}H_{27}O_5$ $[M+H]^+$: 359.1853, found: 359.1858; **IR** (thin film): 2925, 2853, 1717, 1634, 1453 cm^{-1}.



(*R*)-3-((E)-3-methoxy-3-oxoprop-1-en-1-yl)-7-methyl-6,8-dioxo-7,8-dihydro-6H-isochromen-7-yl (2S,4S)-2,4-dimethylhexanoate (3.14)

To a solution of **3.12** (5.5 mg, 0.015 mmol, 1 equiv) and methyl acrylate (5.4 µL, 0.060 mmol, 4.0 equiv) in degassed DCM (0.6 mL) was added a solution of Grubbs Catalyst 2nd generation (2.5 mg, 0.0030 mmol, 0.20 equiv) in degassed DCM (0.5 mL). The mixture was stirred at 45 °C for 4 h before it was cooled to rt and concentrated. Purification on silica gel (10-35% EtOAc in hexanes) afforded 2.6 mg (48% yield) of the title compound as a yellow glass. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.88 (s, 1H), 7.13 (d, J = 15.6 Hz, 1H), 6.52 (d, J = 15.6 Hz, 1H), 6.45 (s, 1H), 5.69 (s, 1H), 3.82 (s, 3H), 2.69 (m, 1H), 1.75 (m,

1H), 1.53 (s, 3H), 1.29 (m, 2H), 1.19 (d, J = 6.8 3H), 1.13 (m, 2H), 0.89 (m, 6H). **^{13}C NMR** (201 MHz, $CDCl_3$) δ 192.8, 192.5, 165.8, 153.2, 153.1, 152.7, 140.7, 133.5, 124.0, 115.8, 114.9, 110.5, 83.8, 52.3, 40.8, 36.2, 31.8, 29.5, 21.9, 19.1, 17.6, 11.1; **HR-ESI-MS**: *m/z* calculated for $C_{22}H_{27}O_6$: 403.1751, found: 403.1728; **IR** (thin film): 2926, 2853, 1718, 1630, 1454 cm^{-1}. $[\alpha]_D$ -12.2 ° (*c* 0.1, $CHCl_3$). All spectra obtained were consistent with literature values.[16]

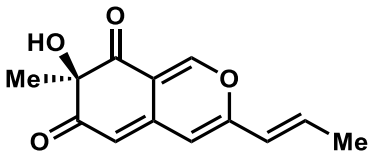
II. Biocatalytic Reactions

Stock solutions: Stock solutions of each substrate (50 mM) were prepared by dissolving the substrate in DMSO (analytical grade). Stock solutions of $NADP^+$ (100 mM) and glucose-6-phosphate (G6P, 500mM) were stored at -20 °C. Aliquots of each flavin-dependent enzyme and glucose-6-phosphate dehydrogenase (G6PDH, 100 U/mL) were stored at -80 °C. Analytical-scale reactions: Each reaction contained 25 µL 100 mM potassium phosphate buffer, pH 8.0, 2.5 mM substrate (2.5 µL of a 50 mM stock solution in DMSO), 5-20 µM flavin-dependent monooxygenase, 5 mM G6P (0.5 µL, 500 mM), 1 mM NADP+ (0.5 µL, 100 mM), 1 U/mL G6P-DH (0.5 µL, 100 U/mL), and Milli-Q water to a final volume of 50 µL. The reaction was carried out at 30 °C for 1 h and quenched by addition of 75 µL acetonitrile with 25 mM pentamethylbenzene as an internal standard. Precipitated biomolecules were pelleted by centrifugation (16,000 x g, 12 min). The supernatant was analyzed by UPLC-DAD and conversion obtained by comparison to calibration curves of each substrate. The subsequent liquid chromatography PDA spectrometry (UPLC) analysis was performed on a Waters Aquity H-Class UPLC-PDA using a Phenomenex Kinetex 1.7 µm C18, 2.1x150 mm column under the following

conditions: Method A: mobile phase (A = deionized water + 0.1% formic acid, B = acetonitrile + 0.1% formic acid), 5% to 100% B over 1.5 min, 100% B for 1.0 min; flow rate, 0.5 mL/min; Method B: mobile phase (A = deionized water + 0.1% formic acid, B = acetonitrile + 0.1% formic acid), 5% to 100% B over 2 min, 100% B for 1 min; flow rate, 0.5 mL/min. Based on calibration curves of the starting materials, the percent conversion of the substrate to dearomatized product was calculated with AUCsubstrate/AUCinternal standard at 270 nm. All reactions were performed and analyzed in triplicate.

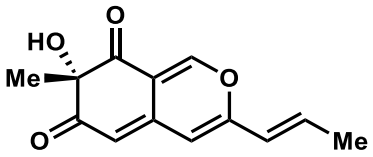
General procedure for *in vitro* preparative-scale reactions: Preparative–scale enzymatic reactions were conducted on 20 mg of each substrate under the following conditions: 5-20 µM flavin-dependent monooxygenase, 2.5 mM substrate, 1 mM NADP+, 1 U/mL G6PDH, and 5 mM G6P for NADPH generation in reaction buffer (50 mM potassium phosphate buffer, pH 8.0). The reaction mixture was added to a 50 mL Erlenmeyer flask and incubated at 30 °C with 100 rpm shaking. After 2 h, a 50 µL aliquot was removed and processed in an identical manner to the analytical-scale reactions described above to determine substrate conversion. The remaining reaction mixture was diluted with acetone (2 x total reaction volume). Precipitated biomolecules were pelleted by centrifugation (4,000 x g, 12 min). Isolation procedure: The supernatant was concentrated under reduced pressure to a final volume of approximately 2 mL. The resulting mixture was filtered through a 0.22 µm filter and purified by preparative HPLC using a Phenomenex Kinetex 5 µm C18, 150 x 21.2 mm column under the following conditions: mobile phase A = deionized water + 0.1% formic acid and B = acetonitrile +

0.1% formic acid; method = 5% to 100% B over 13 min, 100% B for 4 min; flow rate, 15 mL/min.



(*R,E*)-7-hydroxy-7-methyl-3-(prop-1-en-1-yl)-6H-isochromene-6,8(7H)-dione ((*R*)-3.5)

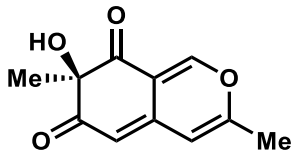
The title compound was synthesized using AzaH according to the general procedure for milligram-scale *in vitro* enzymatic oxidative dearomatization and isolated using the general isolation method. Purification by preparative HPLC afforded 9.6 mg (96% yield) of the title compound as a yellow oil. **^{1}H NMR** (400 MHz, CDl_3) δ 7.89 (s, 1H), 6.59 (m, 1H), 6.10 (s, 1H), 6.01 (d, J = 15.6 Hz, 1H), 5.57 (s, 1H), 2.62 (s, 2H), 1.94 (d, J = 7.0 Hz, 3H), 1.55 (s, 3H). All spectra obtained were consistent with reported values.[9]



(*S,E*)-7-hydroxy-7-methyl-3-(prop-1-en-1-yl)-6H-isochromene-6,8(7H)-dione ((*S*)-3.5)

The title compound was synthesized using AfoD according to the general procedure for milligram-scale *in vitro* enzymatic oxidative dearomatization and isolated using the general isolation method. Purification by preparative HPLC afforded 4 mg (83% yield) of the title compound as a yellow oil. **^{1}H NMR** (400 MHz, CDl_3) δ 7.89 (s, 1H), 6.59 (m, 1H),

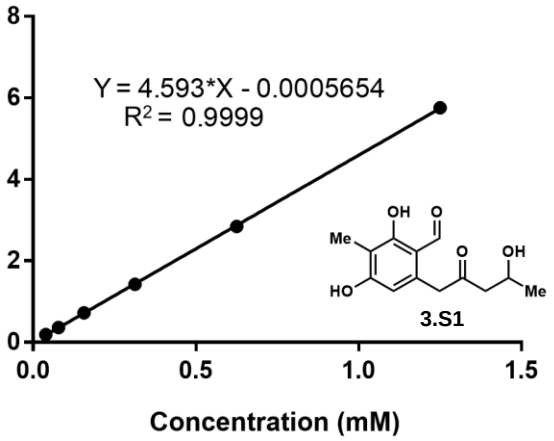
6.10 (s, 1H), 6.01 (d, J = 15.6 Hz, 1H), 5.57 (s, 1H), 2.62 (s, 2H), 1.94 (d, J = 7.0 Hz, 3H), 1.55 (s, 3H). All spectra obtained were consistent with reported values.[9]

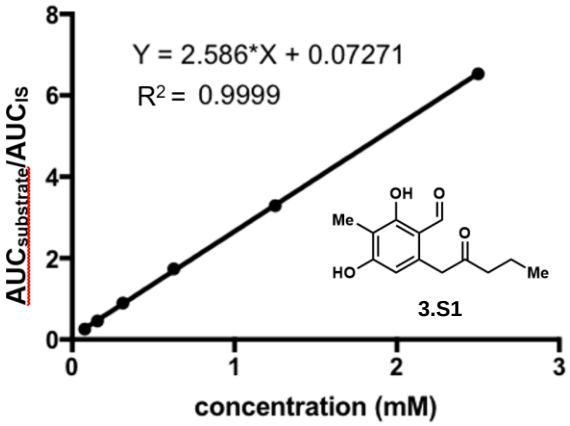


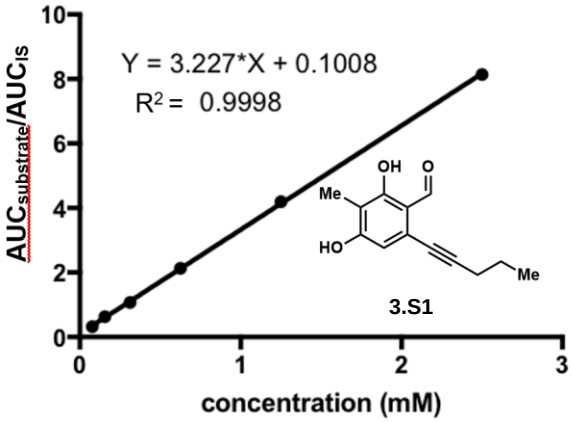
(*R*)-7-hydroxy-3,7-dimethyl-6H-isochromene-6,8(7H)-dione (3.8)

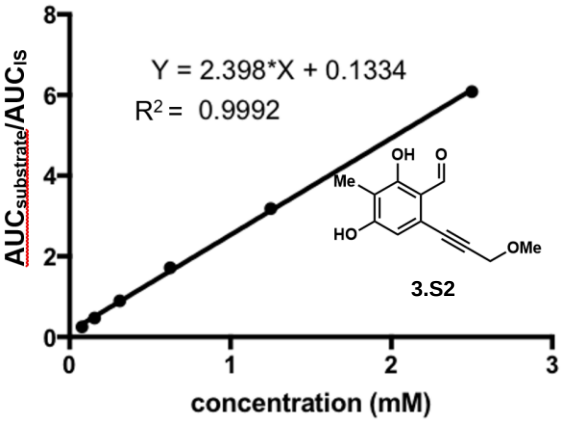
The title compound was synthesized using AzaH according to the general procedure for milligram-scale *in vitro* enzymatic oxidative dearomatization and isolated using the general isolation method. Purification by preparative HPLC afforded 9.5 mg (95% yield) of the title compound as a yellow oil. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.88 (s, 1H), 7.26 (s, 2H), 6.13 (s, 1H), 5.51 (s, 1H), 2.20 (s, 3H), 1.55 (s, 3H). All spectra obtained were consistent with reported values.[4]

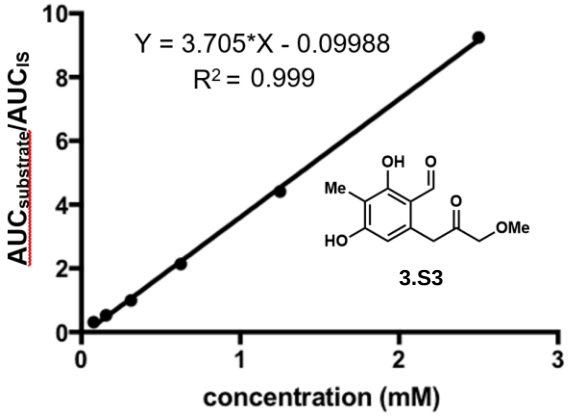
III. Substrate Calibration Curves

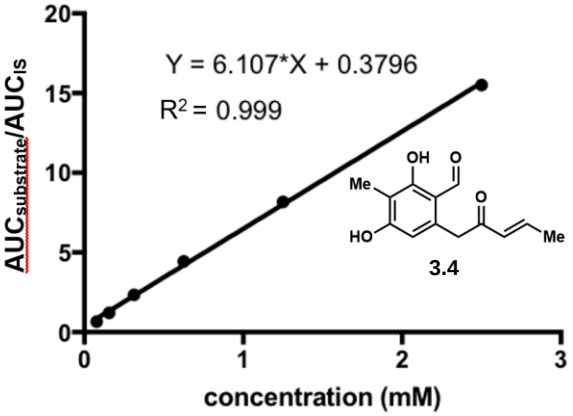


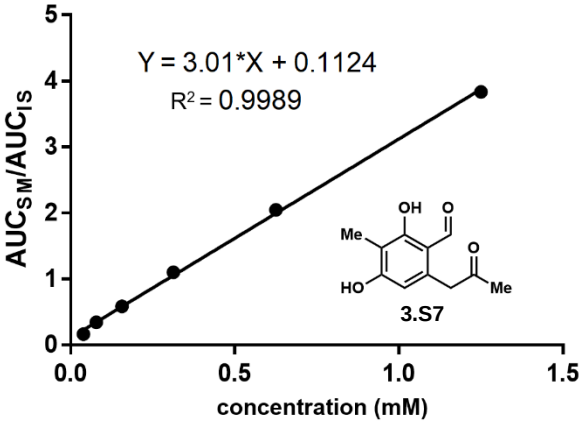








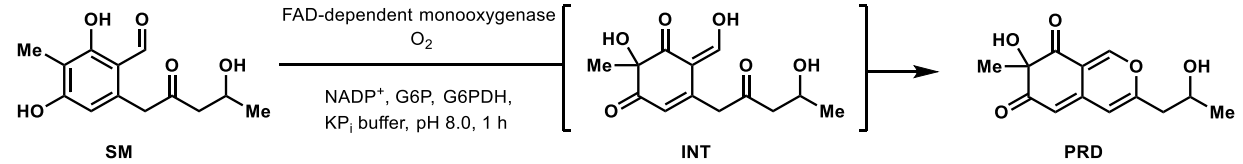



Y = 3.01*X + 0.1124
R2 = 0.9989
AUCSM/AUCIS
concentration (mM)
0.0
0.5
1.0
1.5
0
1
2
3
4
5
OH
O
Me
O
HO
Me

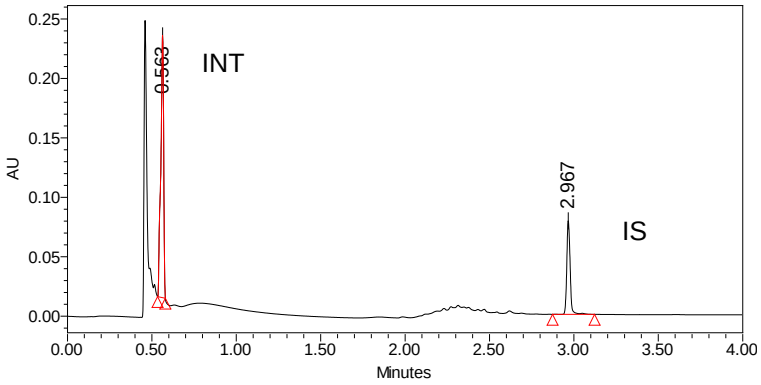
3.S7

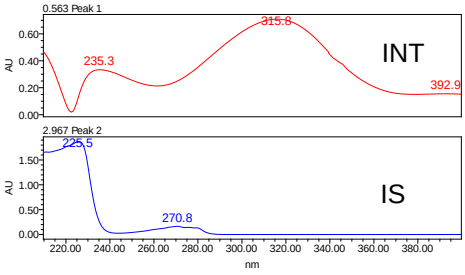
IV. UPLC Traces of Biotransformations

Figure 3.3. Oxidative dearomatization of **3.S12** by AfoD and AzaH. PDA traces of enzymatic reaction and control reaction. (Table 3.1, entry 2). SM = starting material, INT = intermediate, PRD = product, IS = internal standard. The anionic form of the intermediate elutes near the solvent front.



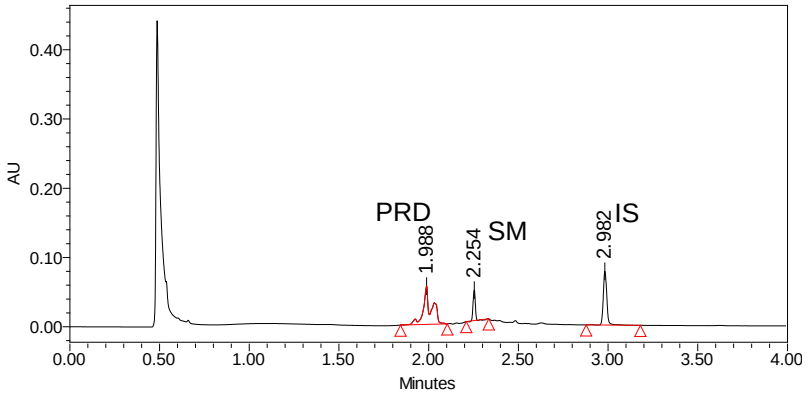
With AzaH

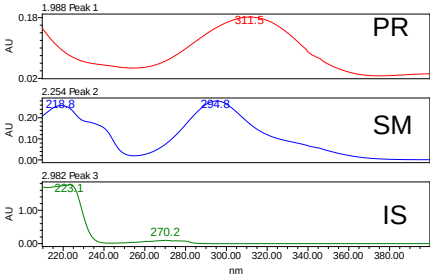




	Retention Time	Area	% Area	Height
1	0.563	258810	70.42	220721
2	2.967	108693	29.58	78693

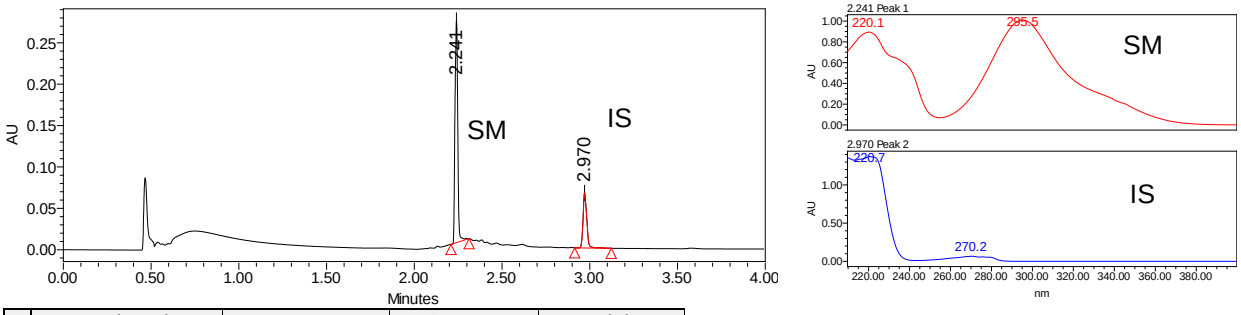
With AfoD





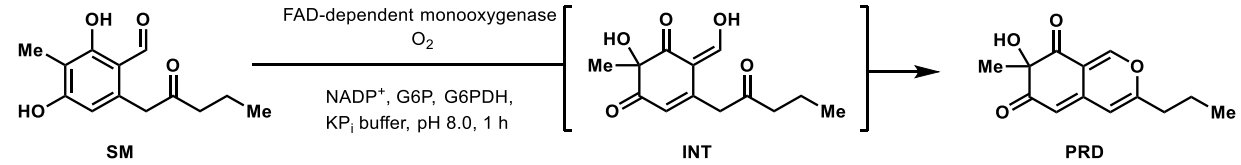
	Retention Time	Area	% Area	Height
1	1.988	174634	54.03	55311
2	2.254	41199	12.75	44106
3	2.982	107399	33.23	77715

NEC

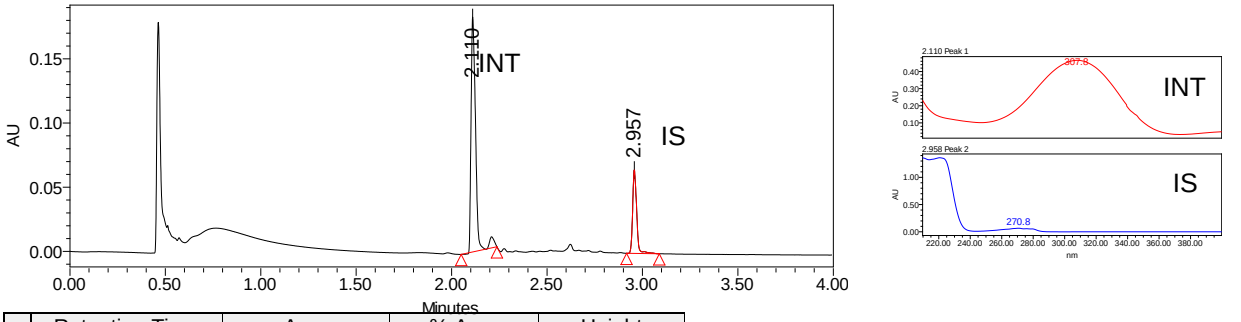


	Retention Time	Area	% Area	Height
1	2.241	292664	74.92	268642
2	2.970	97969	25.08	66721

Figure 3.4. Oxidative dearomatization of **3.S9** by AfoD and AzaH. PDA traces of enzymatic reaction and control reaction (Table 3.1, entry 1).

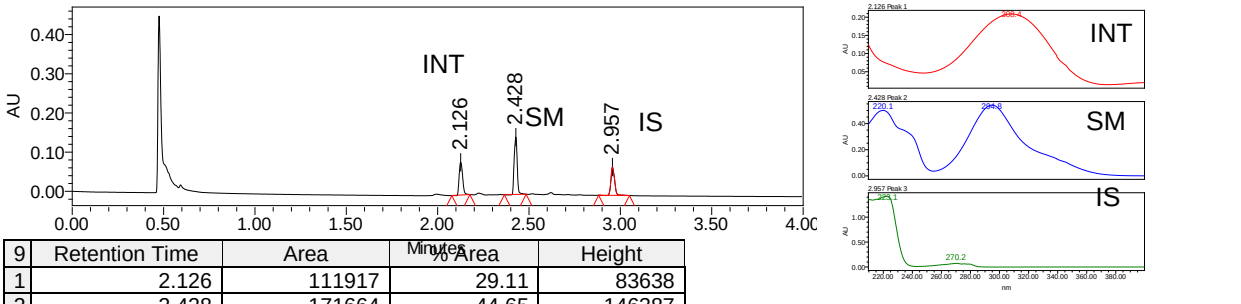


With AzaH



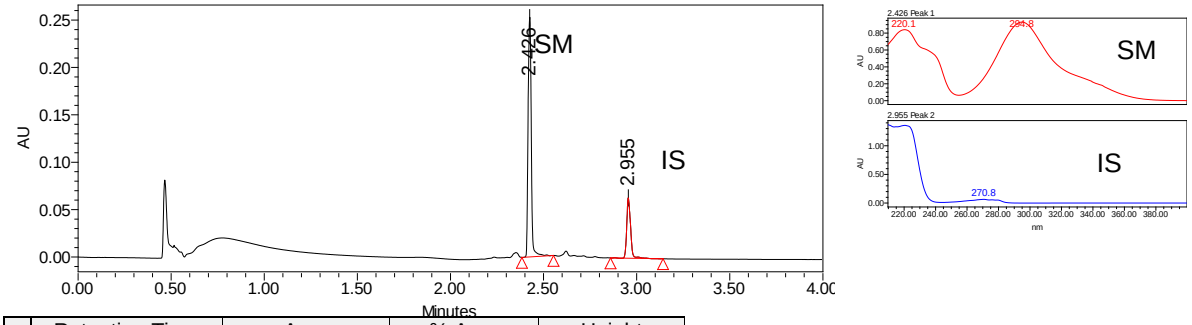
	Retention Time	Area	% Area	Height
1	2.110	305704	76.28	183131
2	2.957	95087	23.72	65154

With AfoD



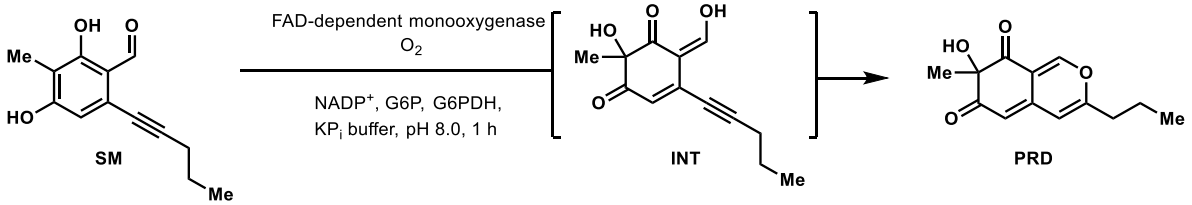
9	Retention Time	Area	% Area	Height
1	2.126	111917	29.11	83638
2	2.428	171664	44.65	146387
3	2.957	100880	26.24	72085

NEC

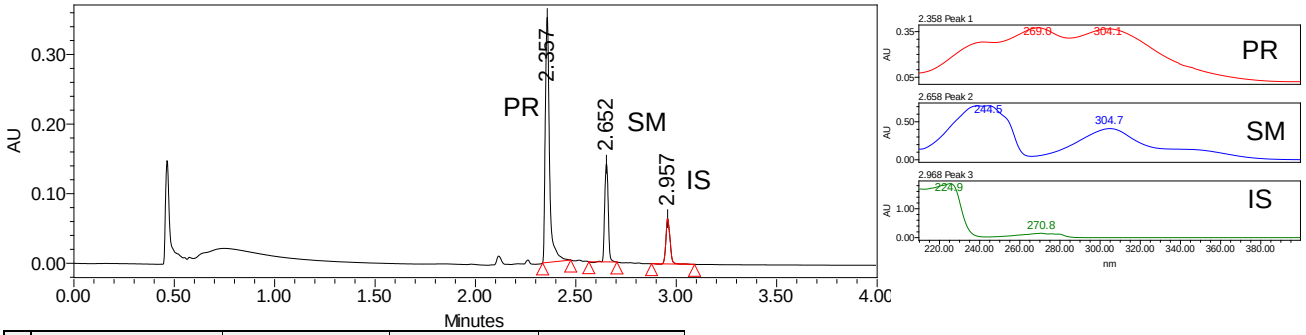


	Retention Time	Area	% Area	Height
1	2.426	301558	76.61	252996
2	2.955	92051	23.39	63755

Figure 3.5. Oxidative dearomatization of **3.S11** by AfoD and AzaH. PDA traces of enzymatic reaction and control reaction. (Table 3.1, entry 3).

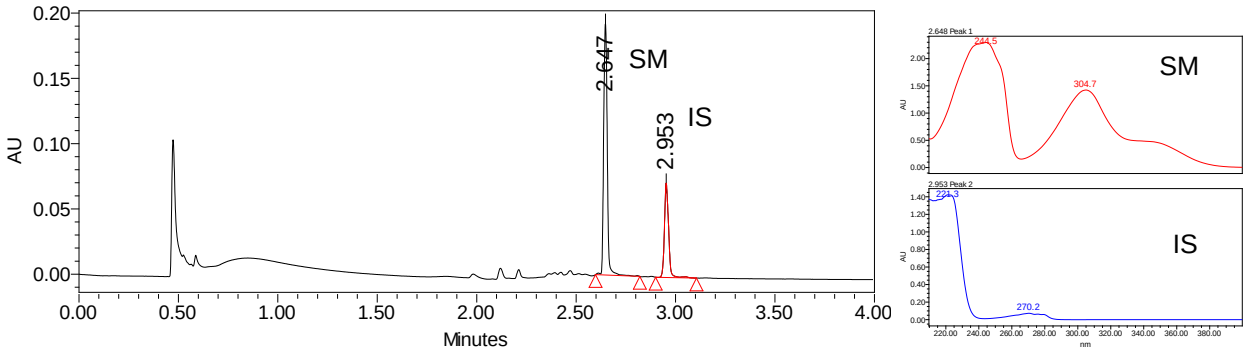


With AzaH



	Retention Time	Area	% Area	Height
1	2.357	517469	65.70	351986
2	2.652	176051	22.35	140720
3	2.957	94126	11.95	65187

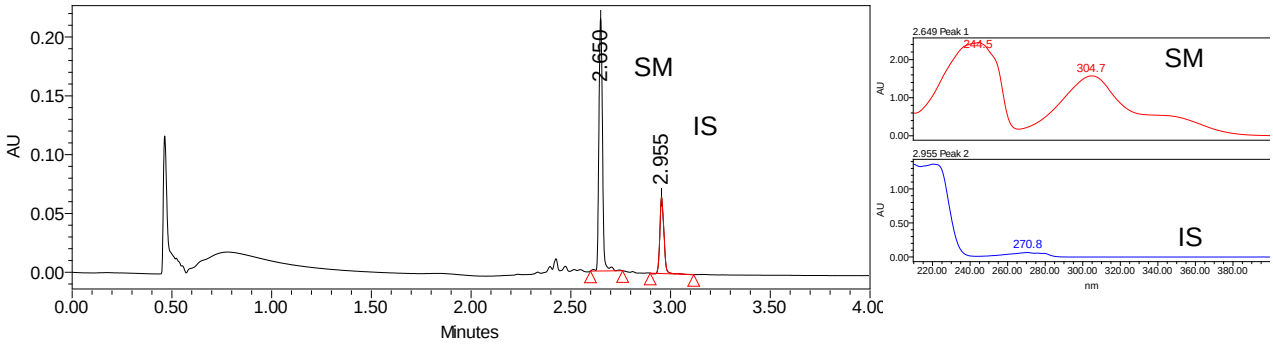
With AfoD



	Retention Time	Area	% Area	Height
1	2.647	240563	69.66	192549

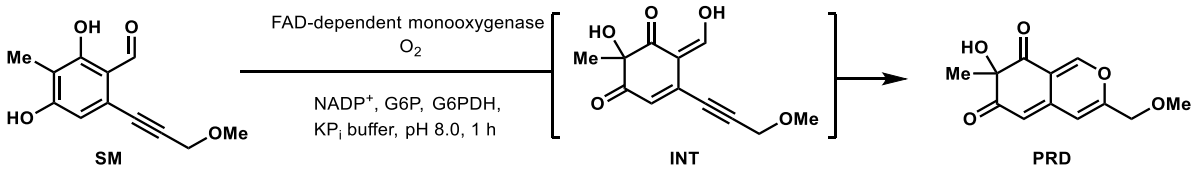
2	2.953	104788	30.34	72317

NEC

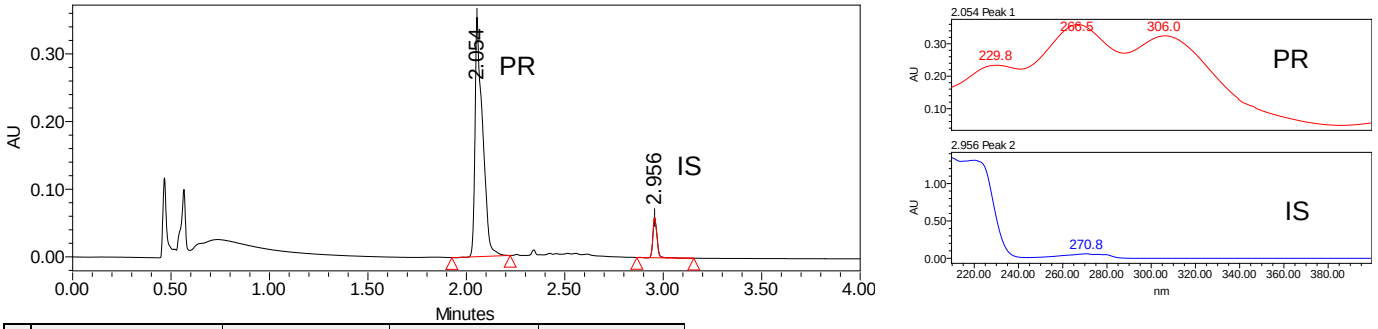


	Retention Time	Area	% Area	Height
1	2.650	266100	74.22	214826
2	2.955	92442	25.78	64989

Figure 3.6. Oxidative dearomatization of **3.S2** by AfoD and AzaH. PDA traces of enzymatic reaction and control reaction. (Table 3.1, entry 4).

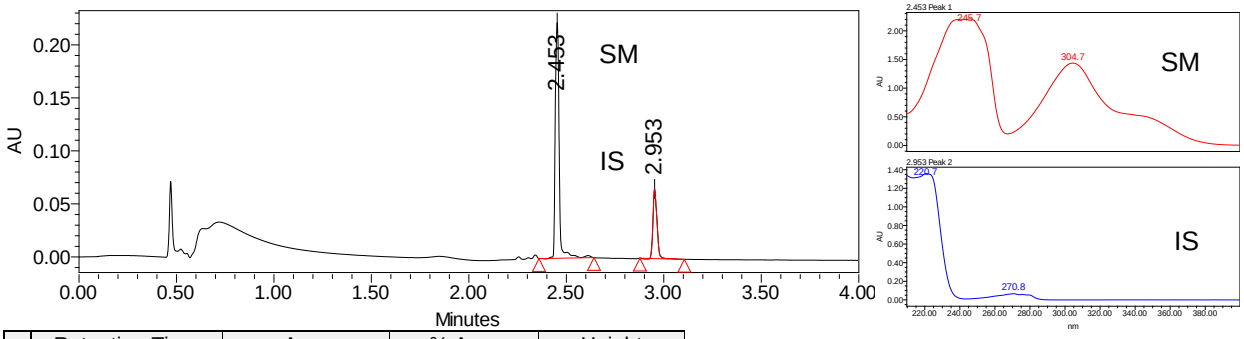


With AzaH



	Retention Time	Area	% Area	Height
1	2.054	944560	91.63	354150
2	2.956	86293	8.37	59688

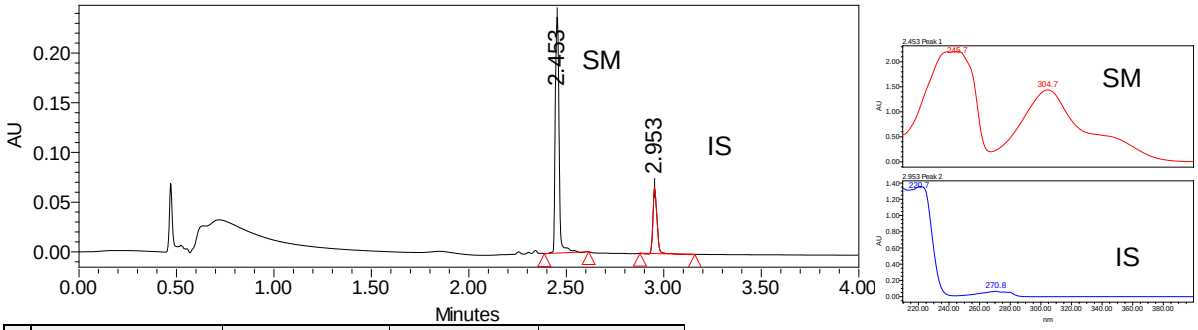
With AfoD



	Retention Time	Area	% Area	Height
1	2.453	309943	76.49	222375

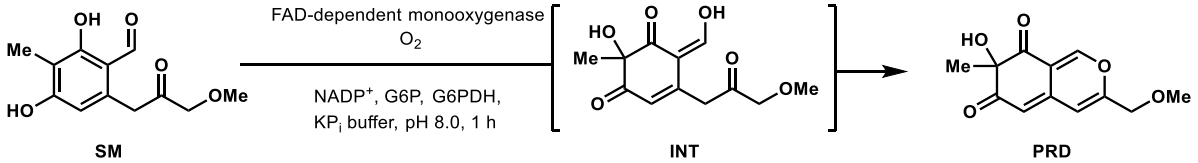
2	2.953	95246	23.51	65655

NEC

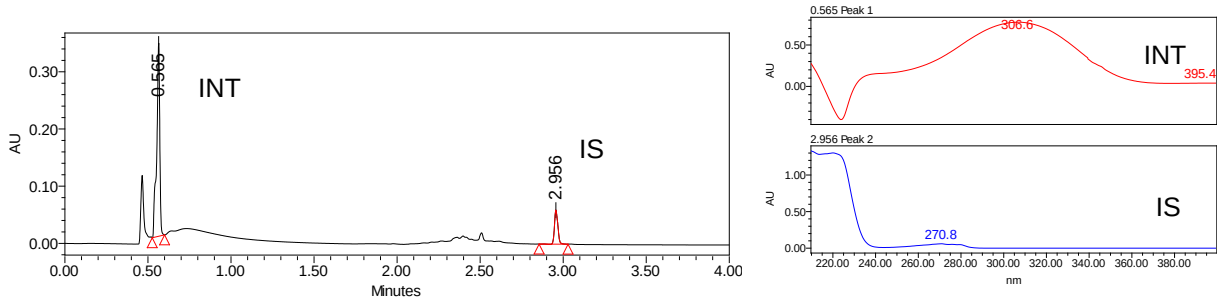


	Retention Time	Area	% Area	Height
1	2.453	321740	77.09	237334
2	2.953	95619	22.91	65898

Figure 3.7. Oxidative dearomatization of **3.S3** by AfoD and AzaH. PDA traces of enzymatic reaction and control reaction. (Table 3.1, entry 5).

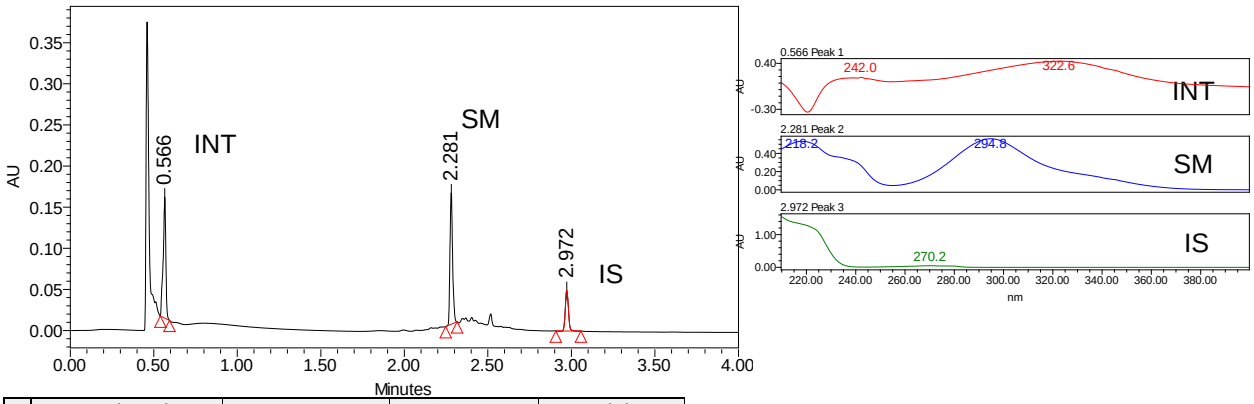


With AzaH



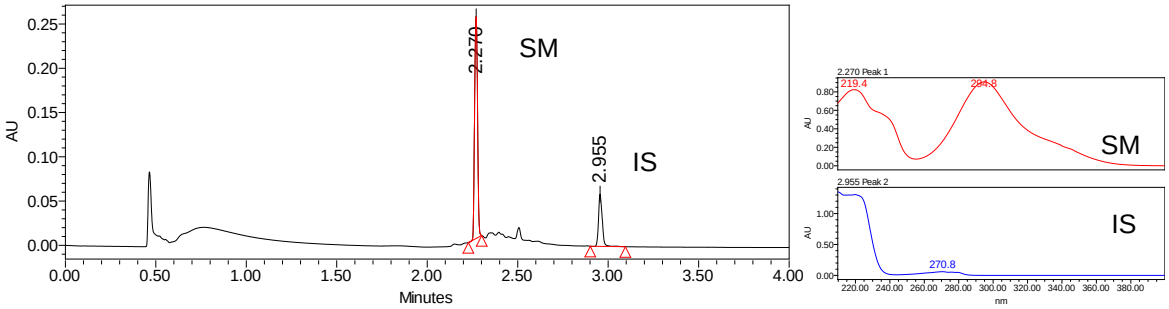
	Retention Time	Area	% Area	Height
1	0.565	434990	83.72	338239
2	2.956	84593	16.28	59668

With AfoD



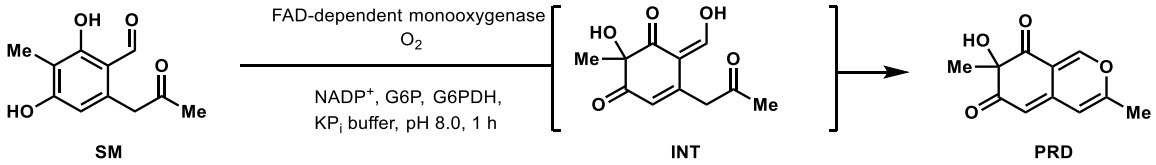
	Retention Time	Area	% Area	Height
1	0.572	113995	30.42	89263
2	2.271	157150	41.94	133701
3	2.953	103585	27.64	72398

NEC

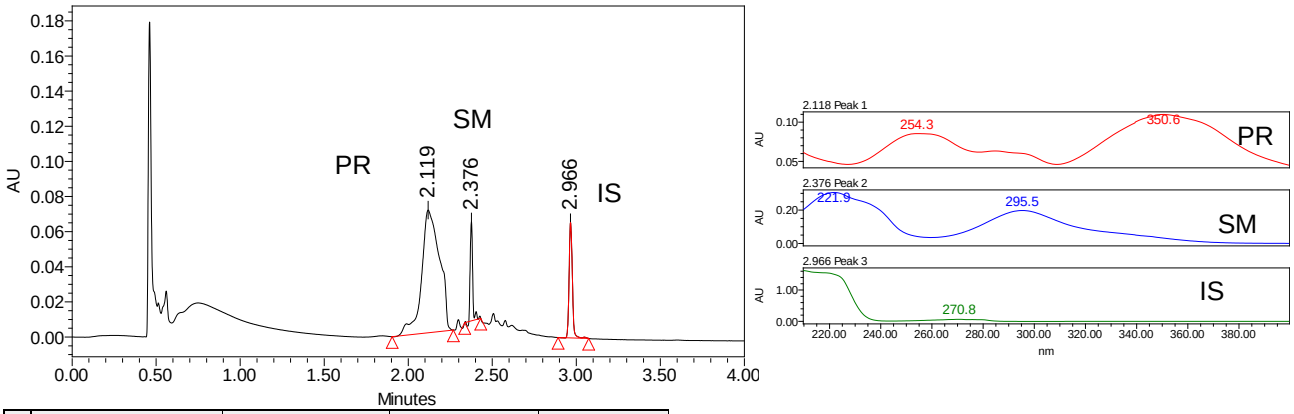


	Retention Time	Area	% Area	Height
1	2.270	284197	77.11	250475
2	2.955	84347	22.89	58999

Figure 3.8. Oxidative dearomatization of **3.7** by AfoD and AzaH. PDA traces of enzymatic reaction and control reaction. (Table 3.1, entry 7).

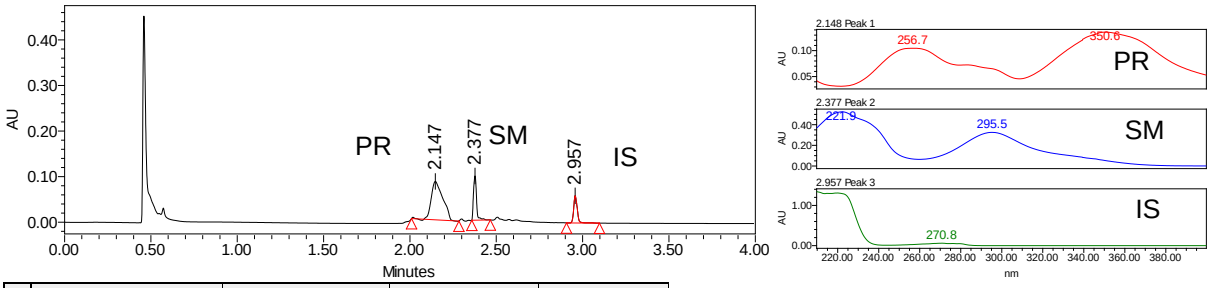


With AzaH



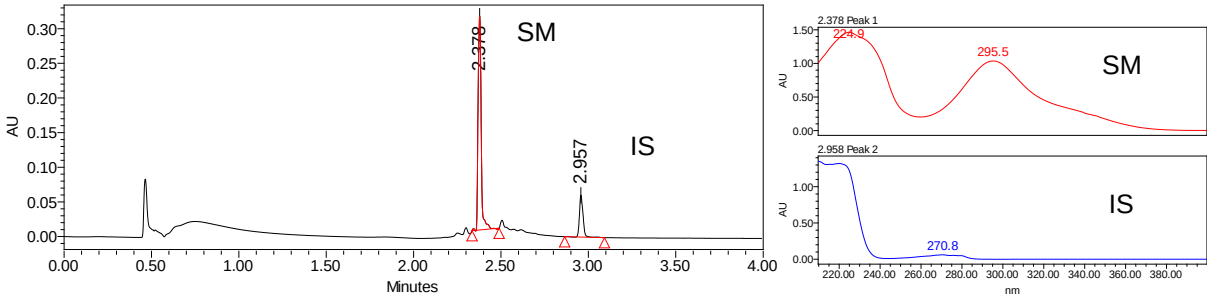
	Retention Time	Area	% Area	Height
1	2.119	519413	77.67	69861
2	2.376	60100	8.99	56370
3	2.966	89232	13.34	65670

With AfoD



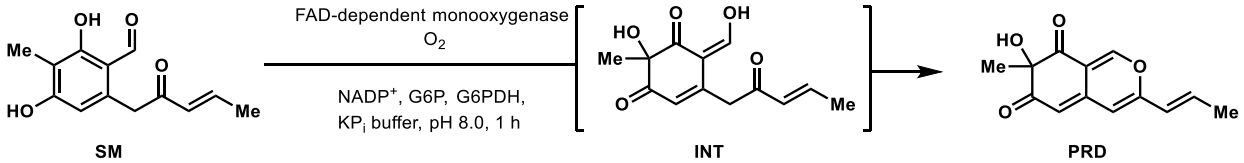
	Retention Time	Area	% Area	Height
1	2.147	398043	65.96	84339
2	2.377	119827	19.86	97170
3	2.957	85588	14.18	59788

NEC

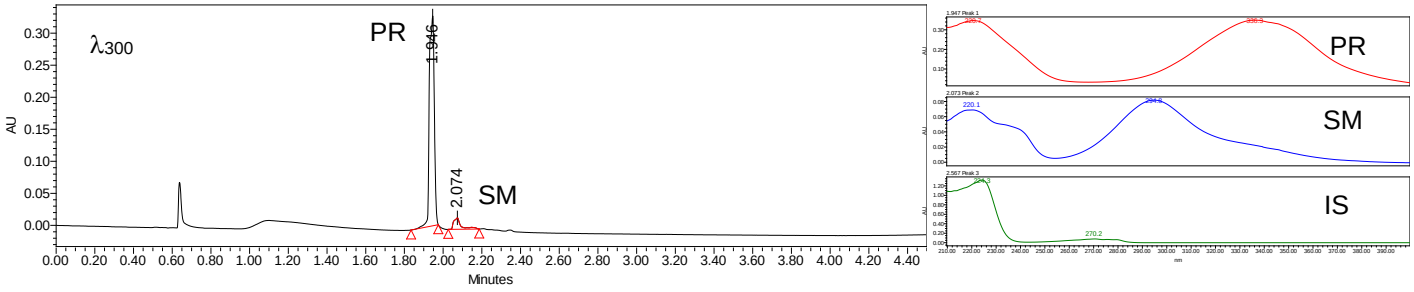


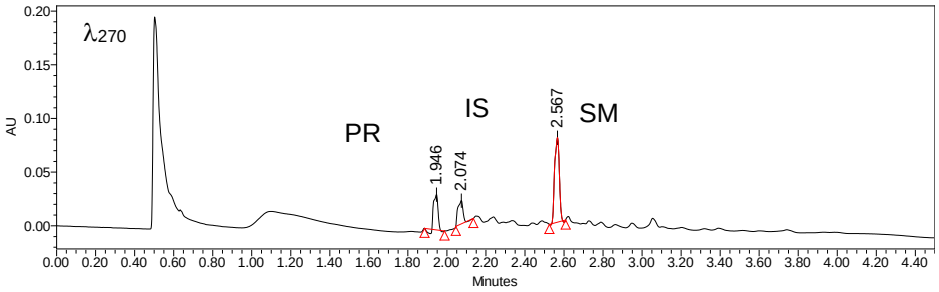
	Retention Time	Area	% Area	Height
1	2.378	383633	81.55	308926
2	2.957	86811	18.45	60584

Figure 3.9. Oxidative dearomatization of **3.4** by AfoD and AzaH. PDA traces of enzymatic reaction and control reaction. (Table 3.1, entry 6). Two wavelengths are shown to visualize product.



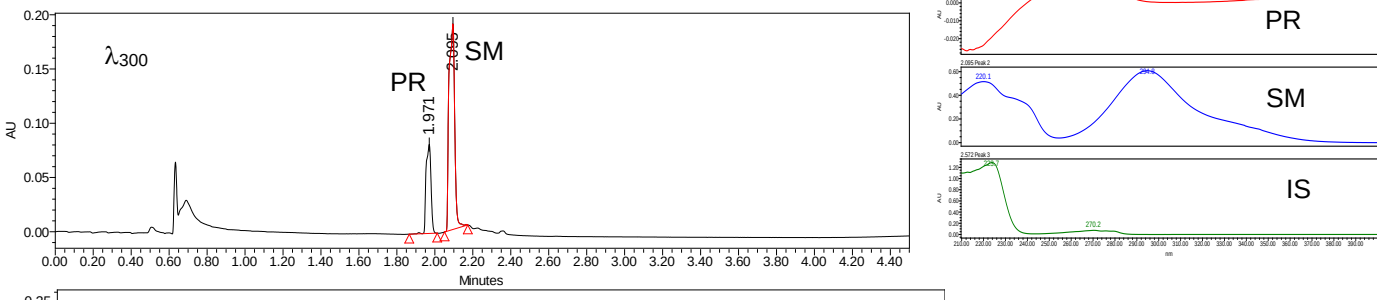
With AzaH

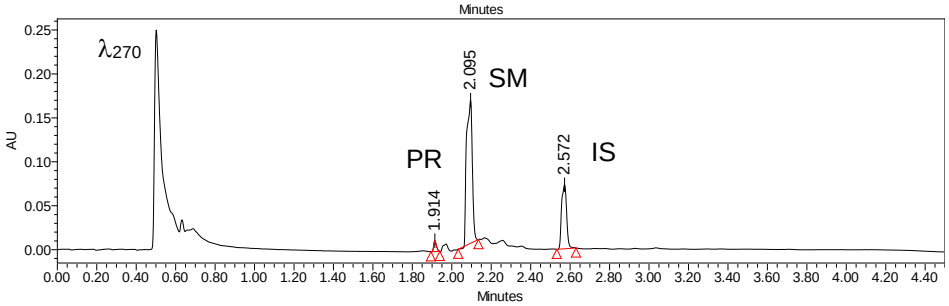




	Retention Time	Area	% Area	Height
1	1.946	59575	25.80	32920
2	2.074	40004	17.33	21701
3	2.567	131297	56.87	78806

With AfoD

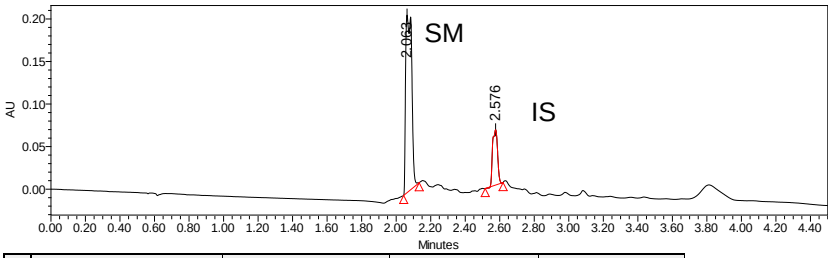




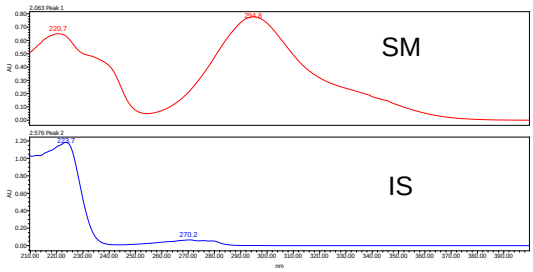
	Retention Time	Area	% Area	Height
1	1.946	59575	25.80	32920
2	2.074	40004	17.33	21701

3	2.567	131297	56.87	78806

NEC



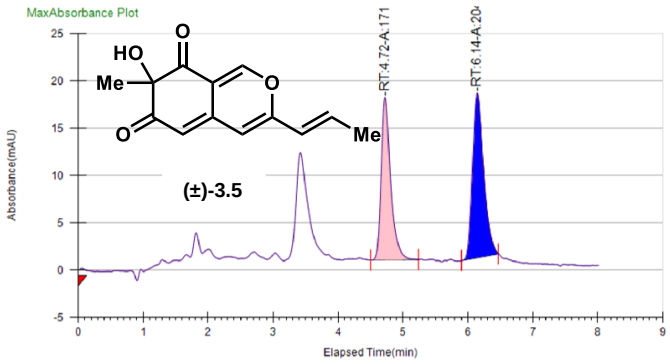
	Retention Time	Area	% Area	Height
1	2.063	507851	79.50	208715
2	2.576	130976	20.50	65081



V. Determination of Enantiomeric Excess

Figure 3.10. PDA traces of racemic **3.5** obtained from a 1:1 mixture of the compound generated using AzaH and AfoD, (*S*)- **3.5** obtained from AfoD-mediated oxidative dearomatization, (*R*)- **3.5** obtained from AzaH-mediated oxidative dearomatization (CHIRALPAK® AD-H, 30%, CO_2, 3.5 mL/min).

Racemic standard



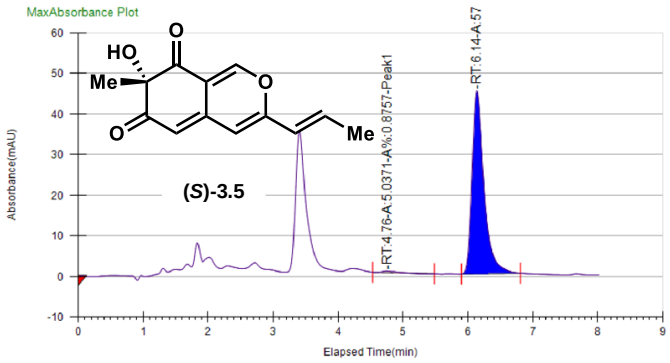
Run Information

Instrument Method	Inj. Vol.	Solvent	Column	Sample	Well Location	Temp.	Flow	% Modifier	Pressure
AD-H_30%_300-330	5	iPrOH	AD-H Chiral Analytical	Rac_bicycle	11A	40	3.5	30	120

Peak Information

Peak No	% Area	Area	Ret. Time	Height	Cap. Factor
1	45.5935	171.6805	4.72 min	17.1822	0
2	54.4065	204.8656	6.14 min	17.4111	0

AfoD reaction



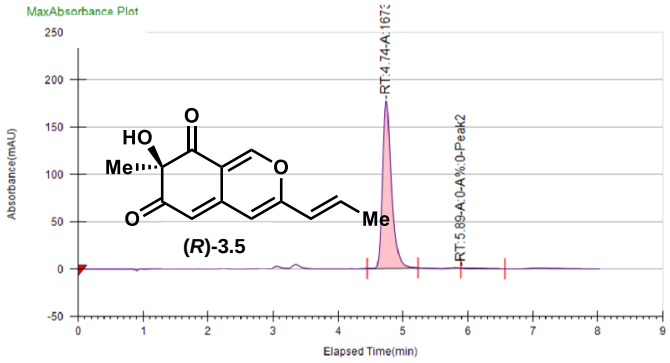
Run Information

Instrument Method	Inj. Vol.	Solvent	Column	Sample	Well Location	Temp.	Flow	% Modifier	Pressure
AD-H_30%_300-330	5	iPrOH	AD-H Chiral Analytical	AfpD_bicycle	13A	40	3.5	30	120

Peak Information

Peak No	% Area	Area	Ret. Time	Height	Cap. Factor
1	0.8757	5.0371	4.76 min	0.3701	0
2	99.1243	570.1434	6.14 min	45.0677	0

AzaH reaction



Run Information

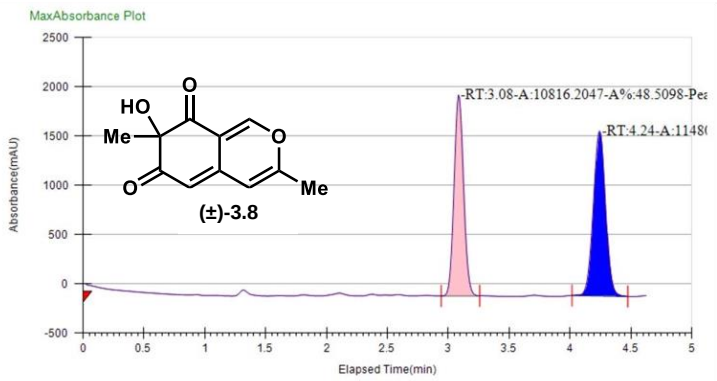
Instrument Method	Inj. Vol.	Solvent	Column	Sample	Well Location	Temp.	Flow	% Modifier	Pressure
AD-H_30%_300-330	5	iPrOH	AD-H Chiral Analytical	AzaH_bicycle	12A	40	3.5	30	120

Peak Information

Peak No	% Area	Area	Ret. Time	Height	Cap. Factor
1	100	1673.5999	4.74 min	176.9264	0
2	0	0	5.89 min	0	0

Figure 3.11. PDA traces of racemic **3.8** obtained from an IBX-mediated oxidative dearomatization, (*S*)- **3.8** obtained from AfoD-mediated oxidative dearomatization, (*R*)-**3.8** obtained from AzaH-mediated oxidative dearomatization (CHIRALPAK® AD-H, 30%, CO_2, 3.5 mL/min).

Racemic standard



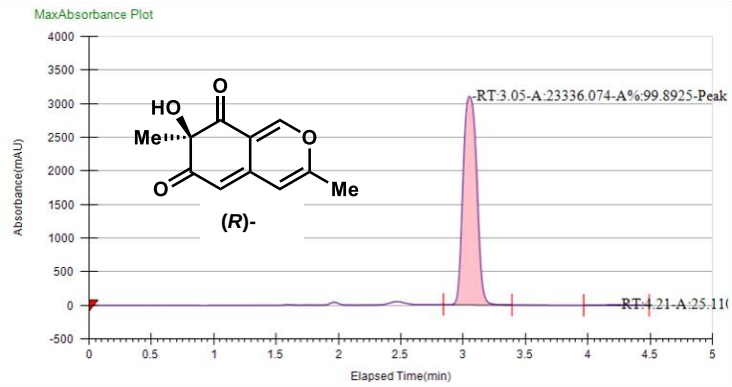
Run Information

Instrument Method	Inj. Vol.	Solvent	Column	Sample	Well Location	Temp.	Flow	% Modifier	Pressure
AD-H_30%_300-330	5	iPrOH	AD-H Chiral Analytical	sbdIV-174_30%	12A	40	3.5	30	120

Peak Information

Peak No	% Area	Area	Ret. Time	Height	Cap. Factor
1	48.5098	10816.2047	3.08 min	2038.3863	3082.3
2	51.4902	11480.7275	4.24 min	1675.0946	4240.6167

AzaH reaction



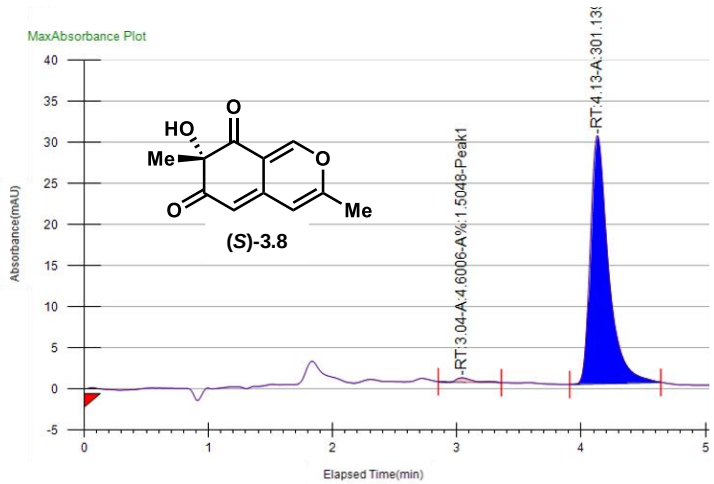
Run Information

Instrument Method	Inj. Vol.	Solvent	Column	Sample	Well Location	Temp.	Flow	% Modifier	Pressure
AD-H_30%_300-330	10	iPrOH	AD-H Chiral Analytical	sbdIV-108_30iPrOH	12A	40	3.5	30	120

Peak Information

Peak No	% Area	Area	Ret. Time	Height	Cap. Factor
1	99.8925	23336.074	3.05 min	3103.1264	3049.0167
2	0.1075	25.1102	4.21 min	2.823	4207.35

AfoD reaction



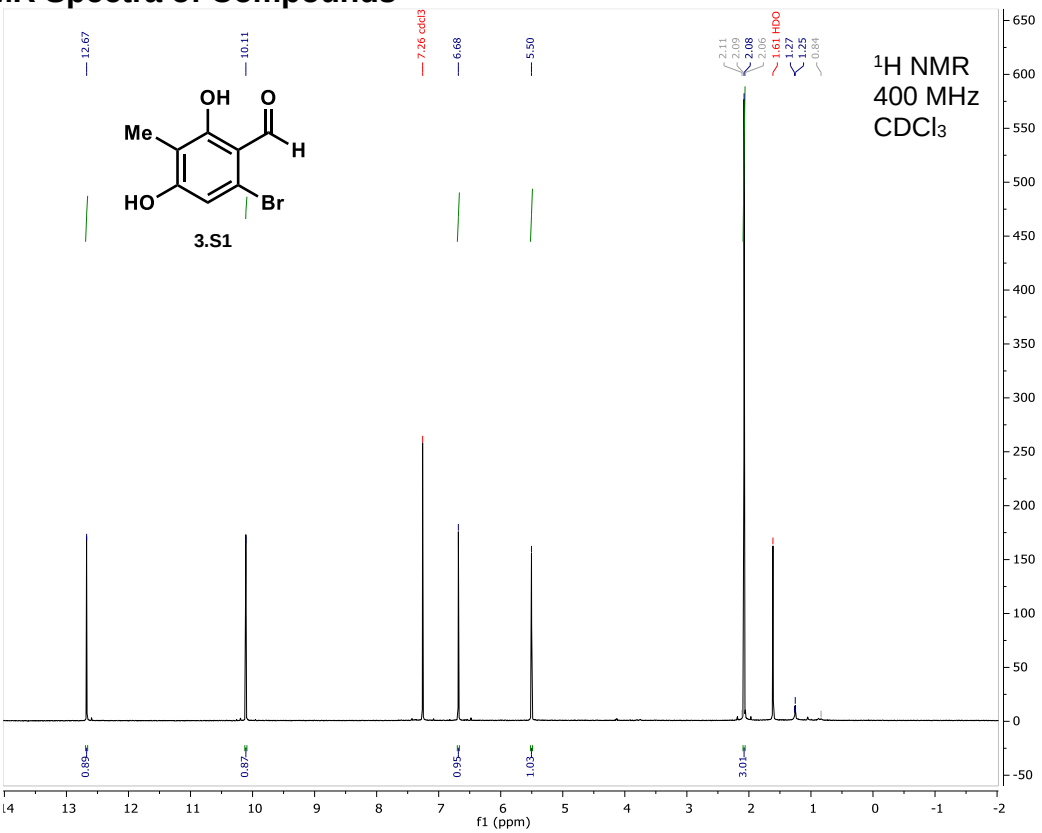
Run Information

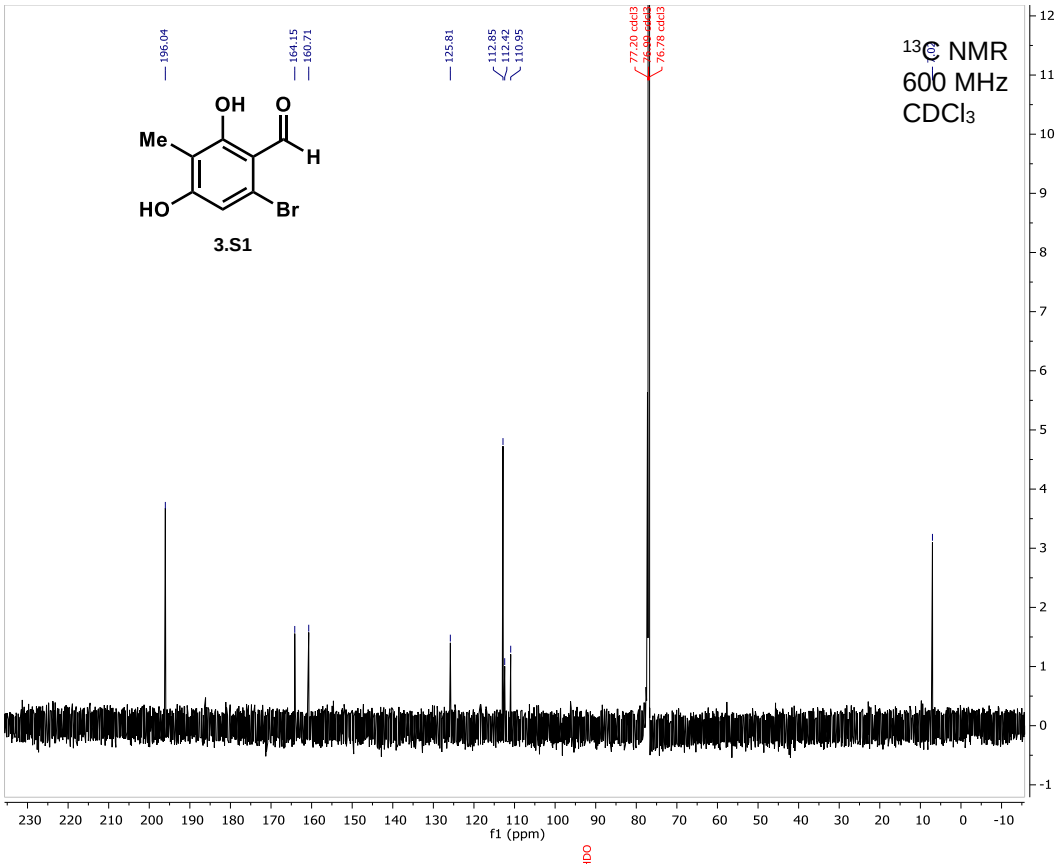
Instrument Method	Inj. Vol.	Solvent	Column	Sample	Well Location	Temp.	Flow	% Modifier	Pressure
AD-H_30%_300-330	10	iPrOH	AD-H Chiral Analytical	JBP-3-169-AfoD	13A	40	3.5	30	120

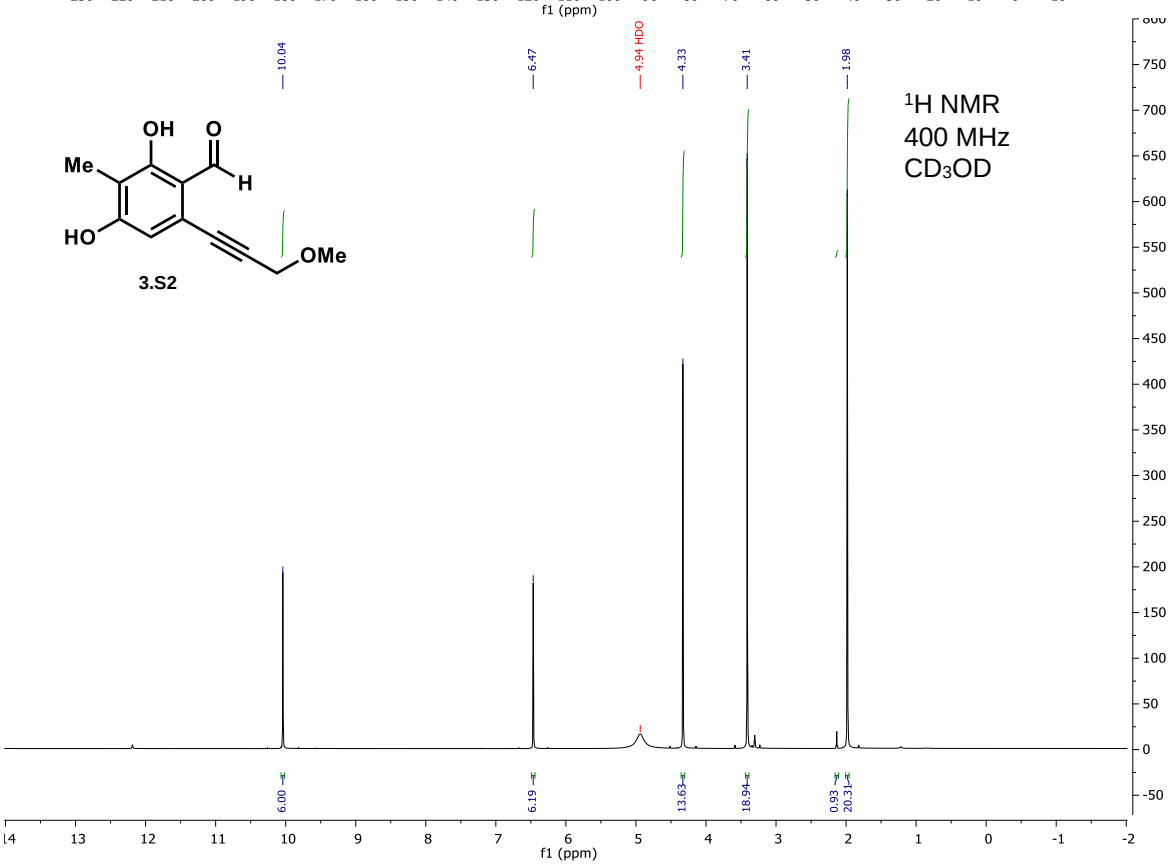
Peak Information

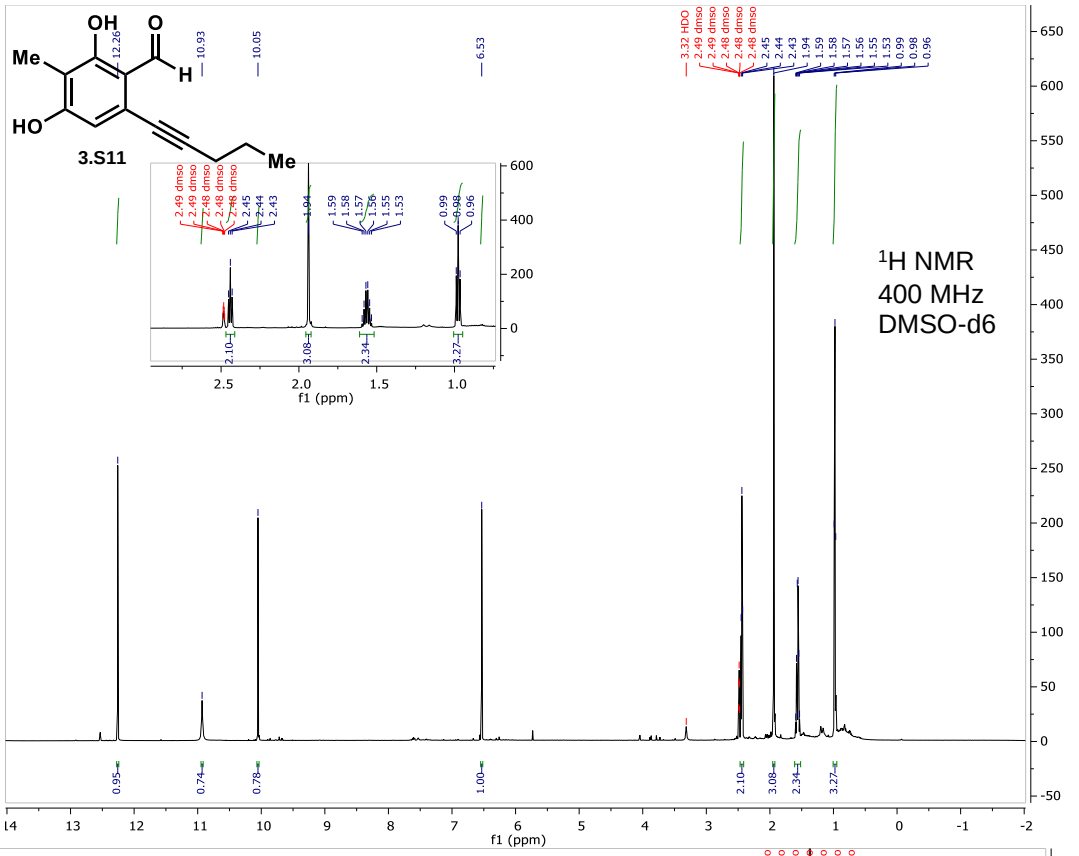
Peak No	% Area	Area	Ret. Time	Height	Cap. Factor
1	1.5048	4.6006	3.04 min	0.51	0
2	98.4952	301.1399	4.13 min	30.1492	0

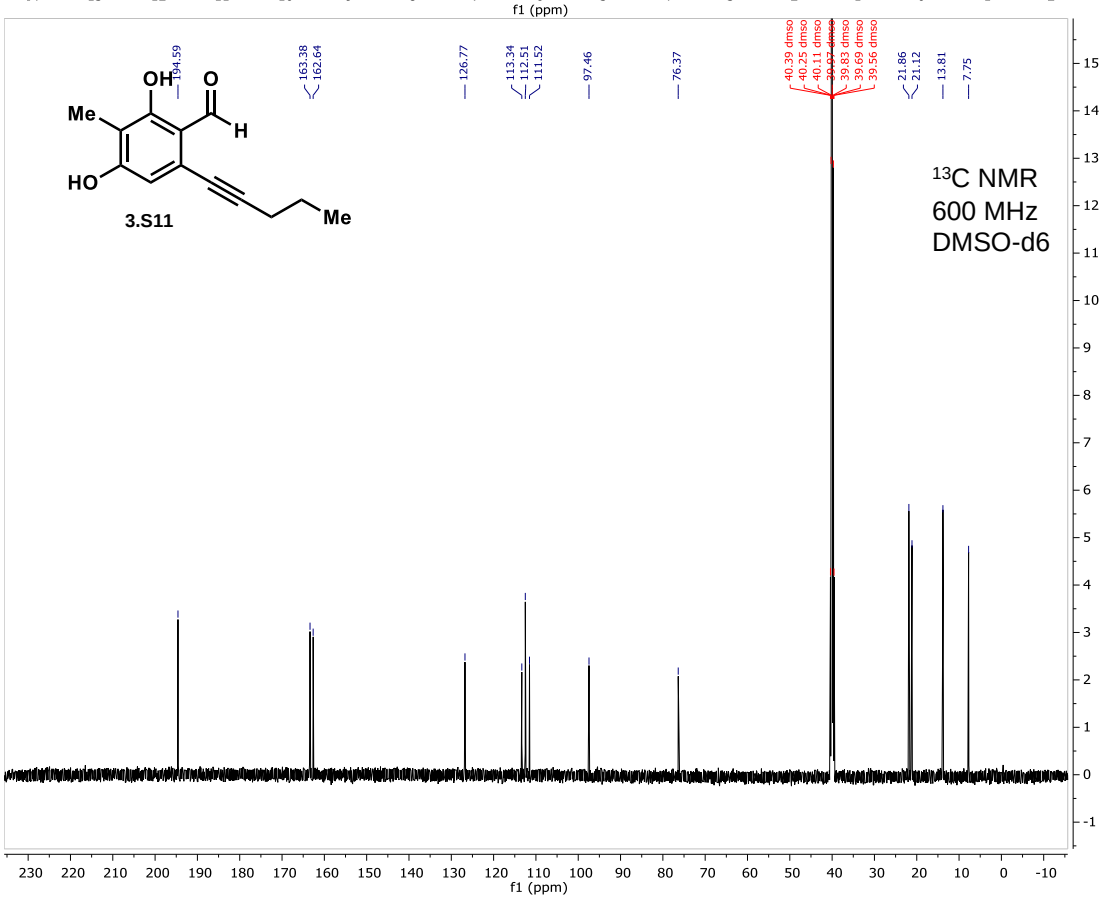
VI. NMR Spectra of Compounds

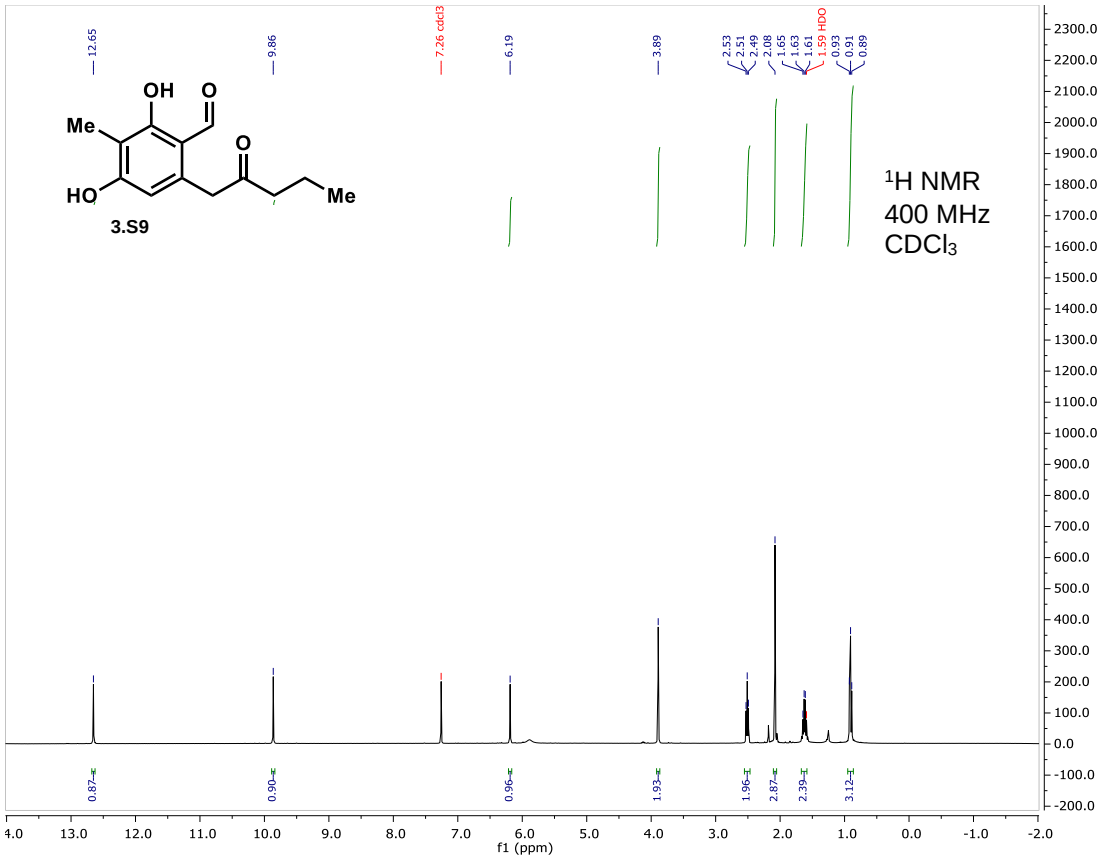


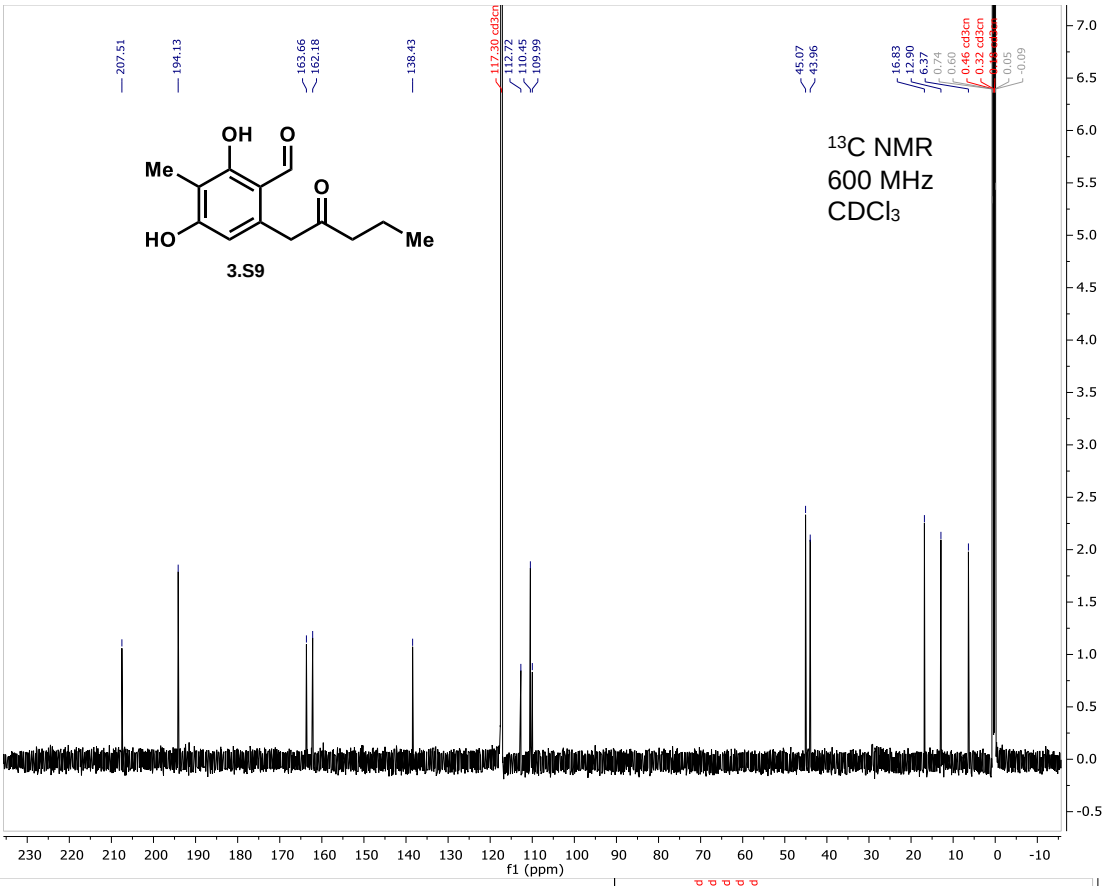

196.04
164.15
160.71
125.81
112.85
112.42
110.95
77.20 cdcl3
76.78 cdcl3
OH
O
Me
H
HO
Br
3.S1
13C NMR
600 MHz
CDCl3
f1 (ppm)

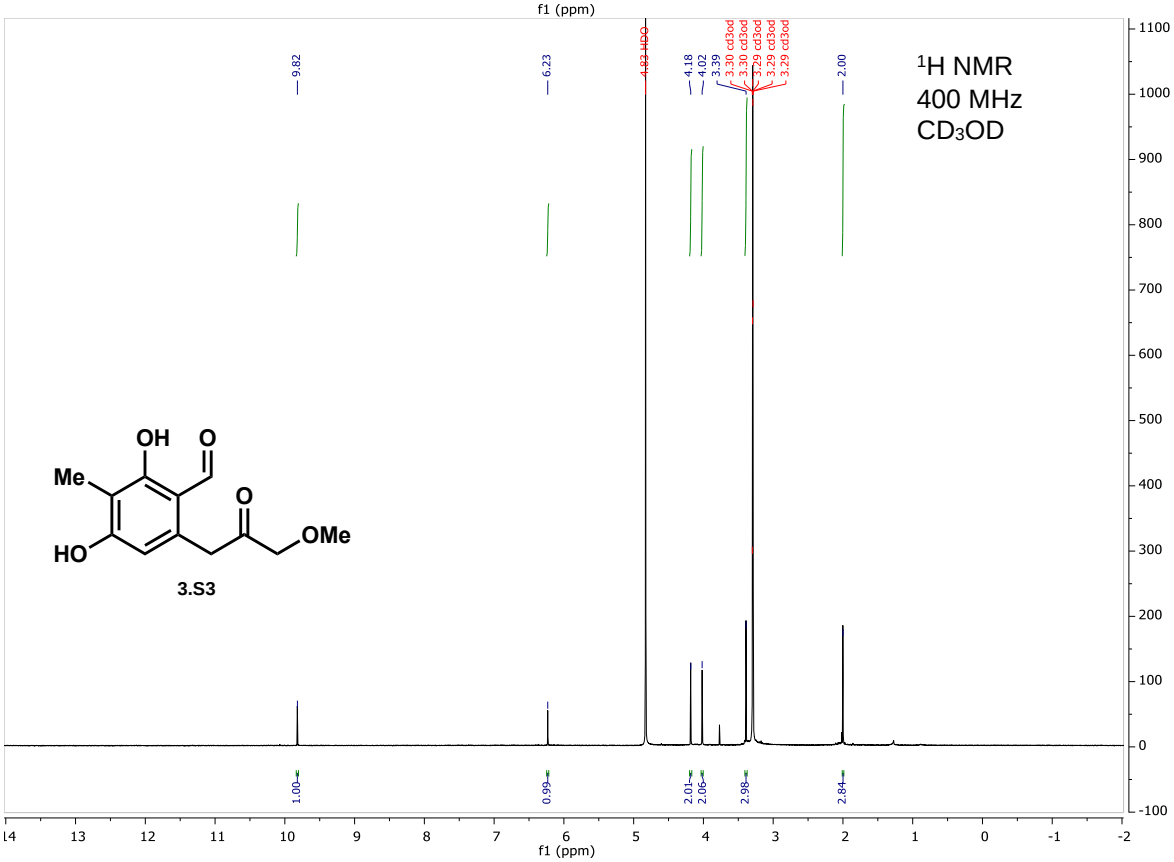

10.04
6.47
4.94 HDO
4.33
3.41
1.98
OH
O
Me
H
HO
OMe
3.S2
1H NMR
400 MHz
CD3OD
6.00
6.19
13.63
18.94
0.93
20.31
f1 (ppm)

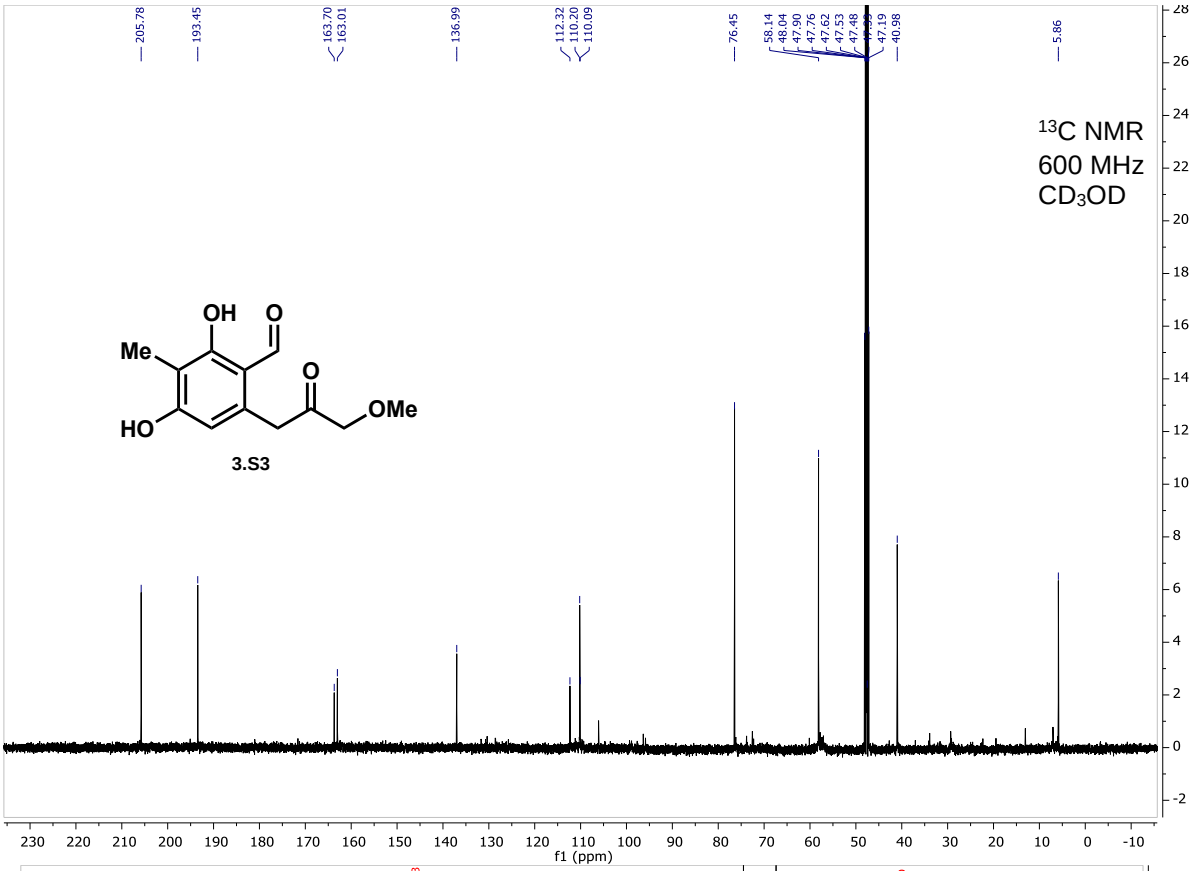

OH
O
Me
H
HO
3.S11
Me
^{1}H NMR
400 MHz
DMSO-d6
f1 (ppm)

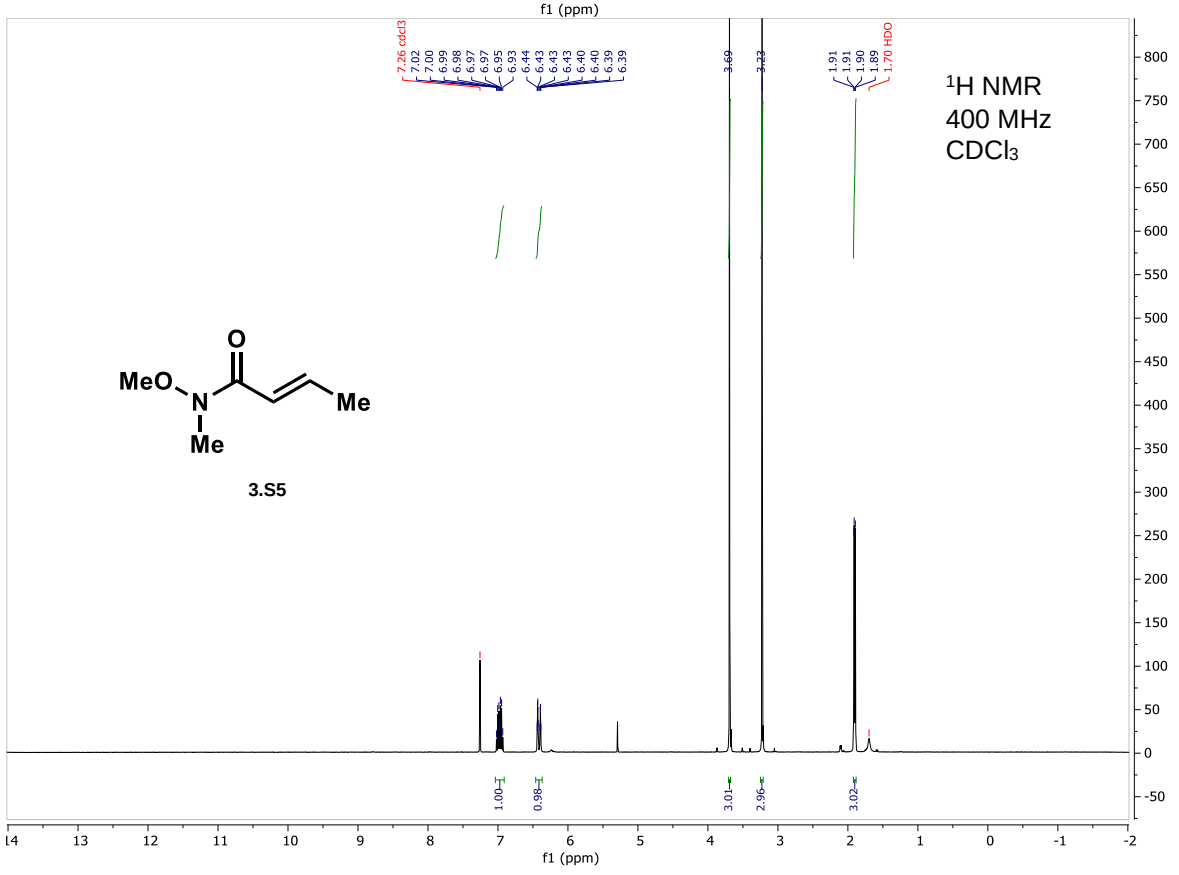

OH
O
Me
H
HO
3.S11
Me
^{13}C NMR
600 MHz
DMSO-d6
f1 (ppm)

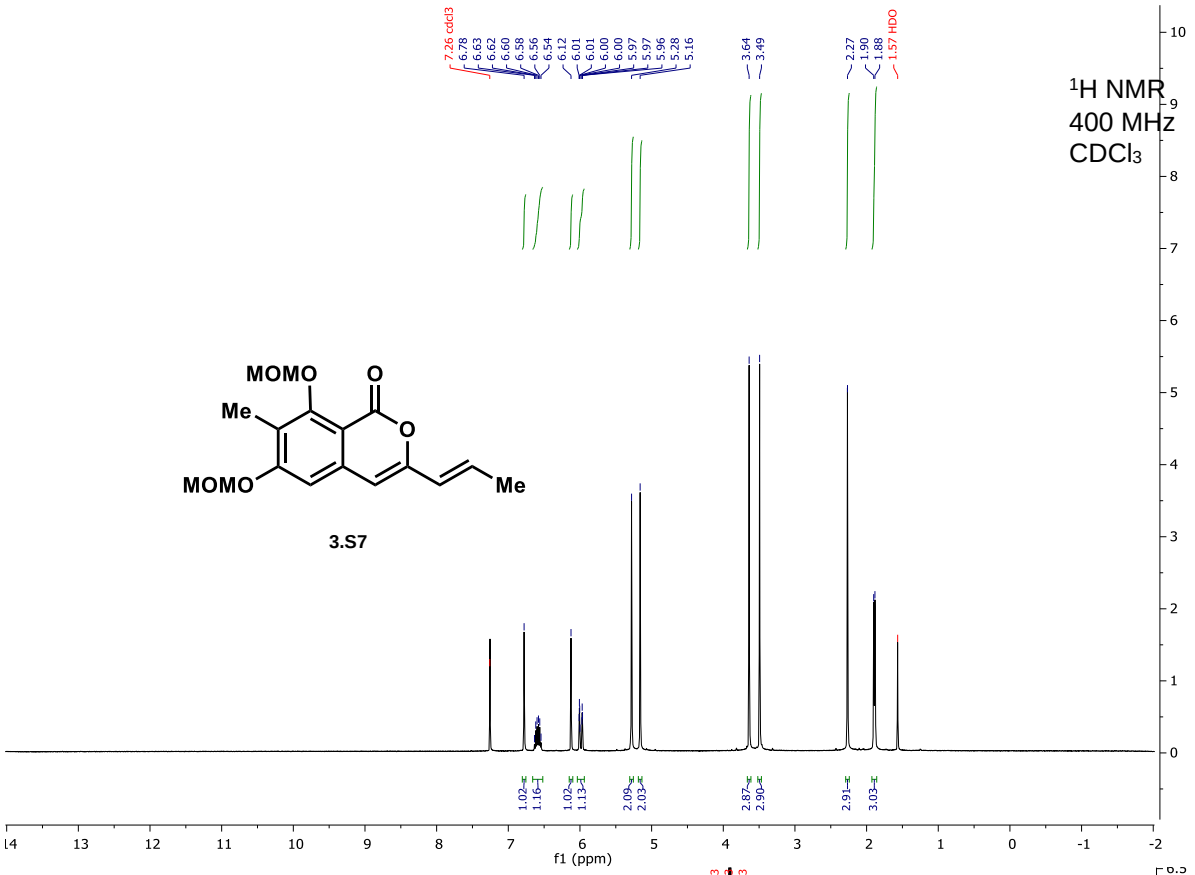

12.65
9.86
7.26 cdcl3
6.19
3.89
2.53
2.51
2.49
2.08
1.65
1.63
1.61
1.59 HDO
0.93
0.91
0.89
OH
O
Me
O
HO
Me
3.S9
1H NMR
400 MHz
CDCl3
0.87
0.90
0.96
1.93
1.96
2.87
2.39
3.12
f1 (ppm)

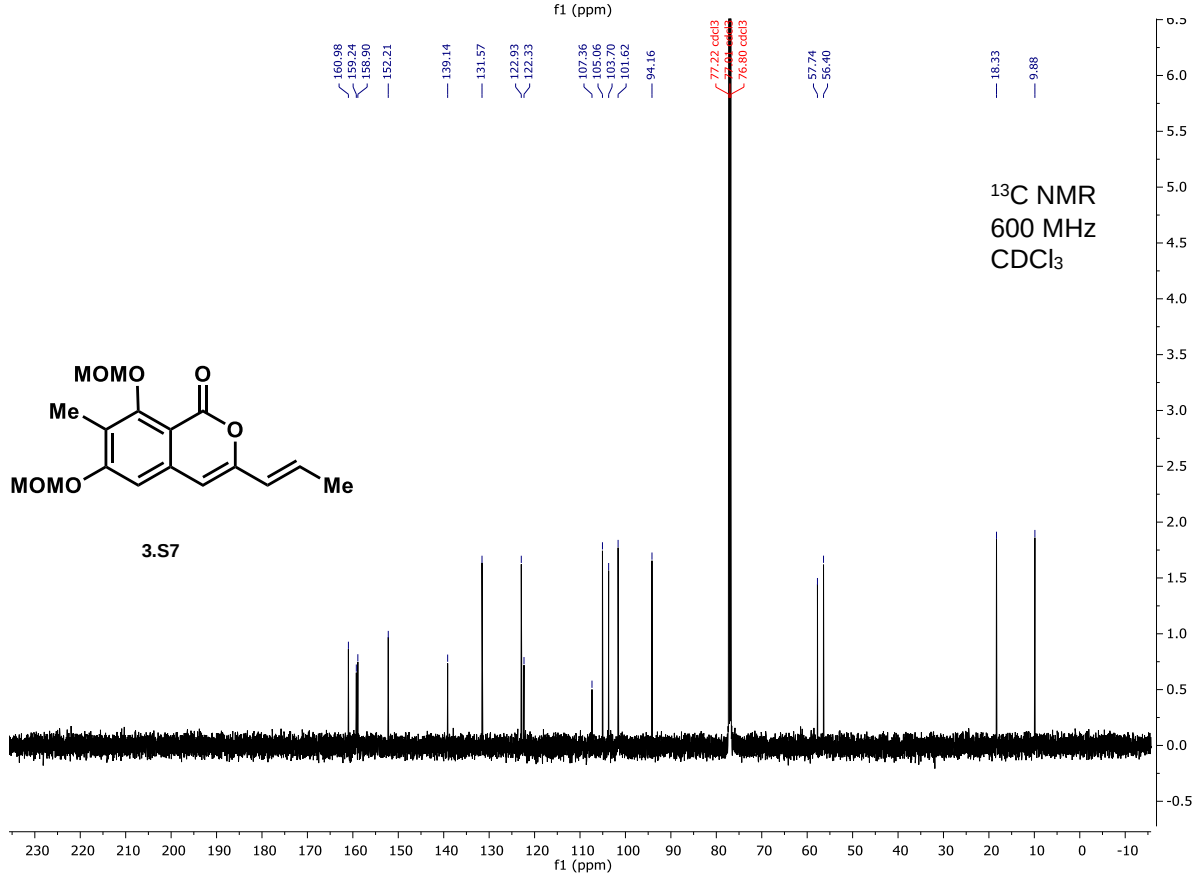

207.51
194.13
163.66
162.18
138.43
117.30 cd3cn
112.72
110.45
109.99
45.07
43.96
16.83
12.90
6.37
0.74
0.60
0.46 cd3cn
0.32 cd3cn
0.05
-0.09
OH
O
Me
O
HO
Me
3.S9
13C NMR
600 MHz
CDCl3
f1 (ppm)

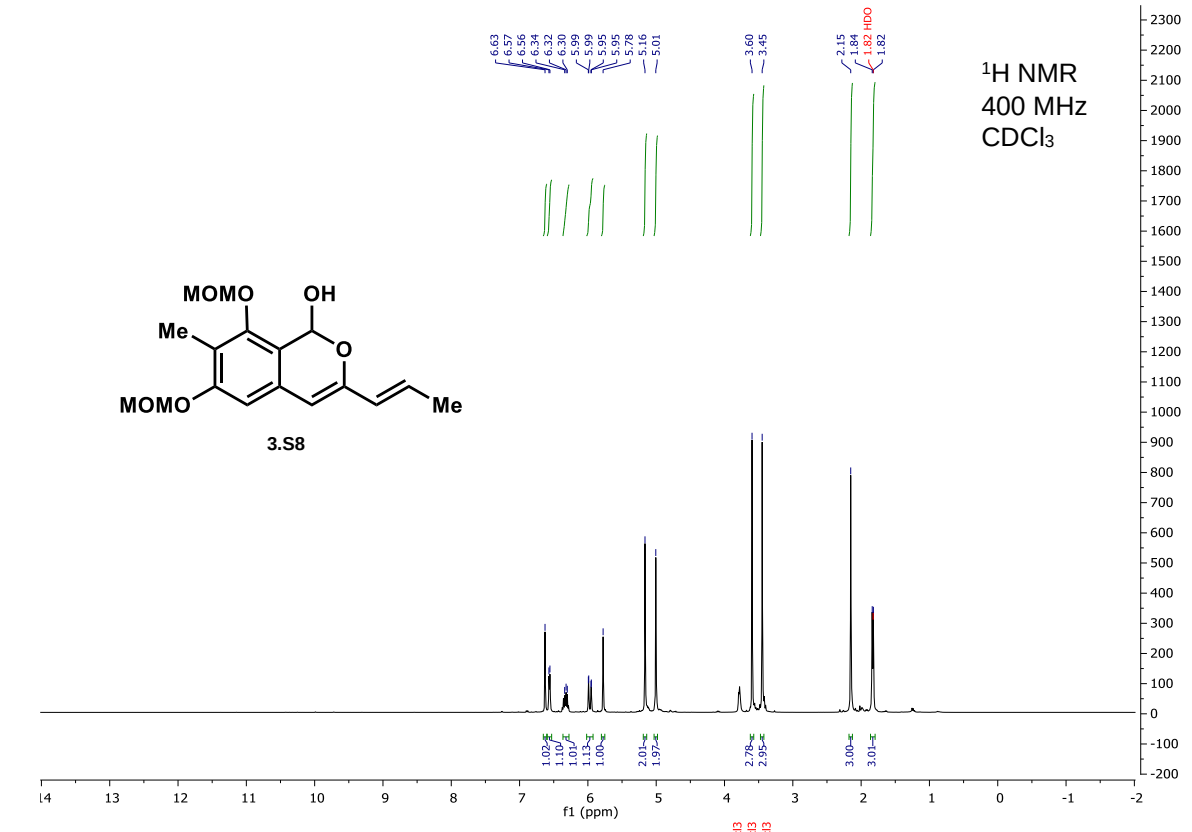

9.82
6.23
4.83 HDO
4.18
4.02
3.39
3.30 cd3od
3.30 cd3od
3.29 cd3od
3.29 cd3od
3.29 cd3od
2.00
1.00
0.99
2.01
2.06
2.98
2.84
OH
O
Me
O
HO
OMe
3.S3
1H NMR
400 MHz
CD3OD
f1 (ppm)

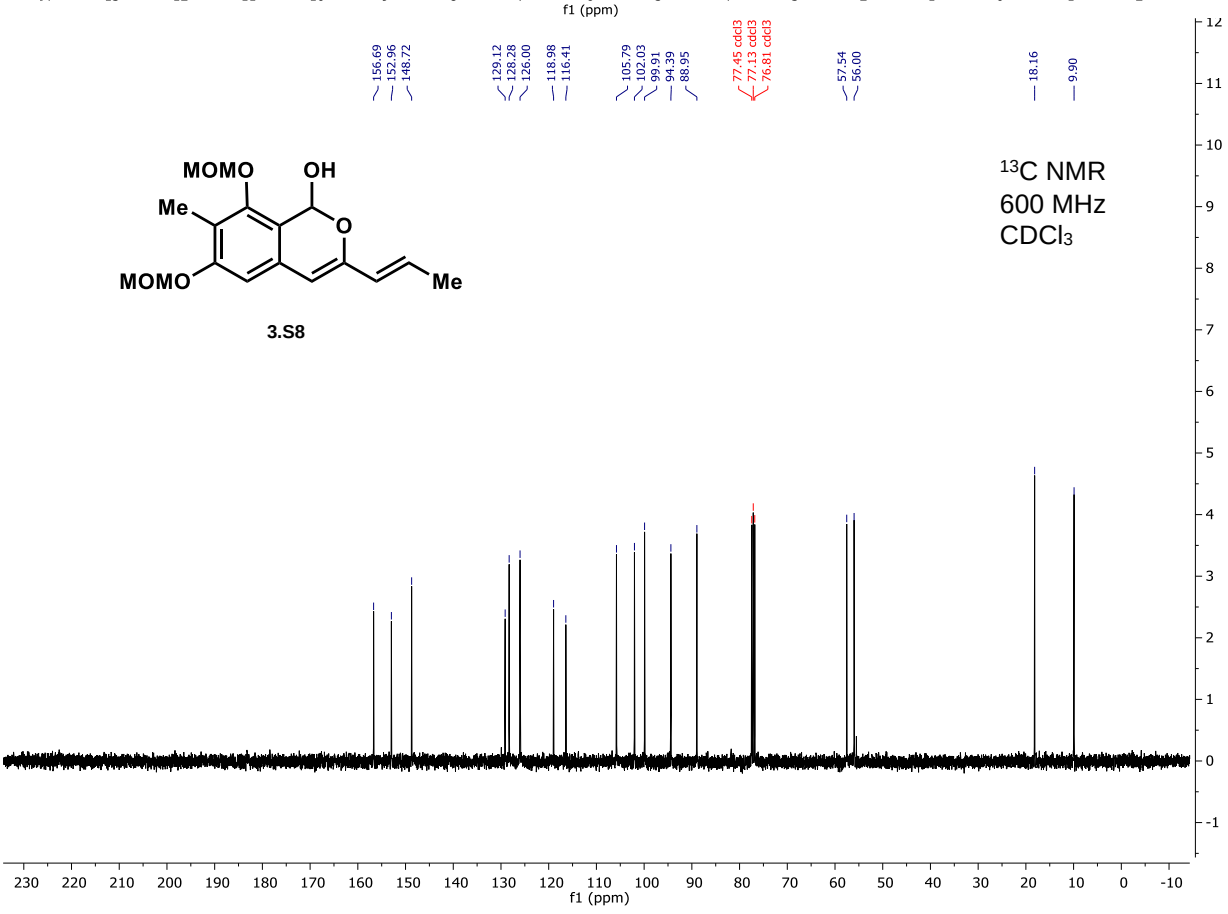

205.78
193.45
163.70
163.01
136.99
112.32
110.20
110.09
76.45
58.14
48.04
47.90
47.76
47.62
47.53
47.48
47.19
40.98
5.86
^{13}C NMR
600 MHz
CD_3OD
OH
O
Me
O
OMe
HO
3.S3
f1 (ppm)

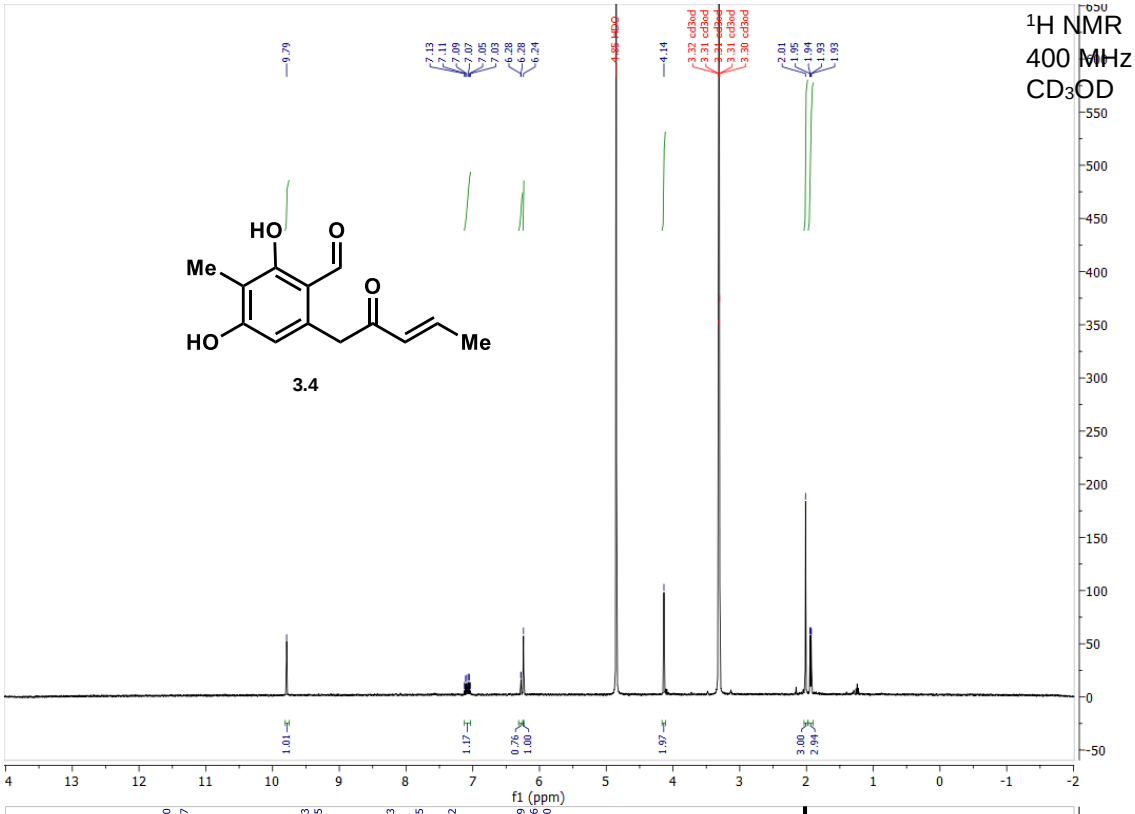

7.26 cdcl3
7.02
7.00
6.99
6.98
6.97
6.97
6.95
6.93
6.44
6.43
6.43
6.43
6.40
6.40
6.39
6.39
3.69
3.23
1.91
1.91
1.90
1.89
1.70 HDO
^{1}H NMR
400 MHz
$CDCl_3$
MeO
N
Me
O
Me
Me
3.S5
1.00
0.98
3.01
2.96
3.02
f1 (ppm)

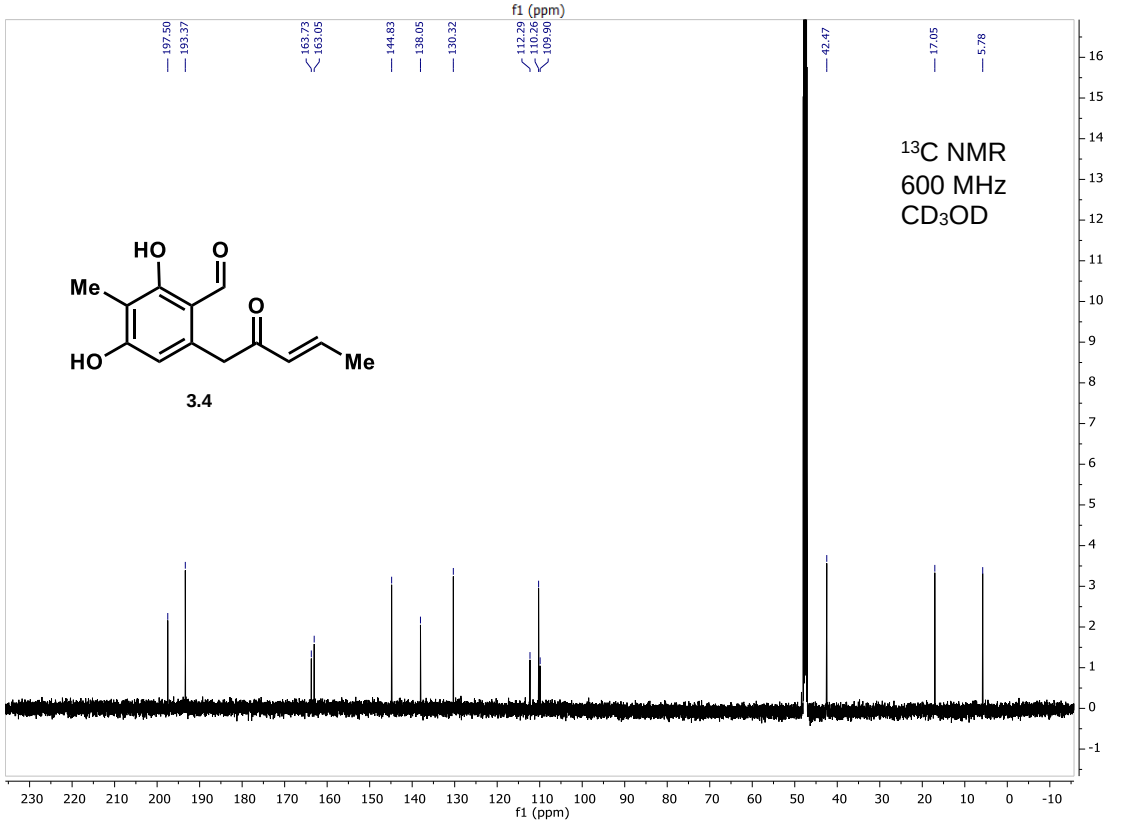

7.26 cdcl3
6.78
6.63
6.62
6.60
6.58
6.56
6.54
6.12
6.01
6.01
6.00
6.00
5.97
5.97
5.96
5.28
5.16
3.64
3.49
2.27
1.90
1.88
1.57 HDO
1H NMR
400 MHz
CDCl3
MOMO
O
Me
O
MOMO
Me
3.S7
1.02
1.16
1.02
1.13
2.09
2.03
2.87
2.90
2.91
3.03
f1 (ppm)

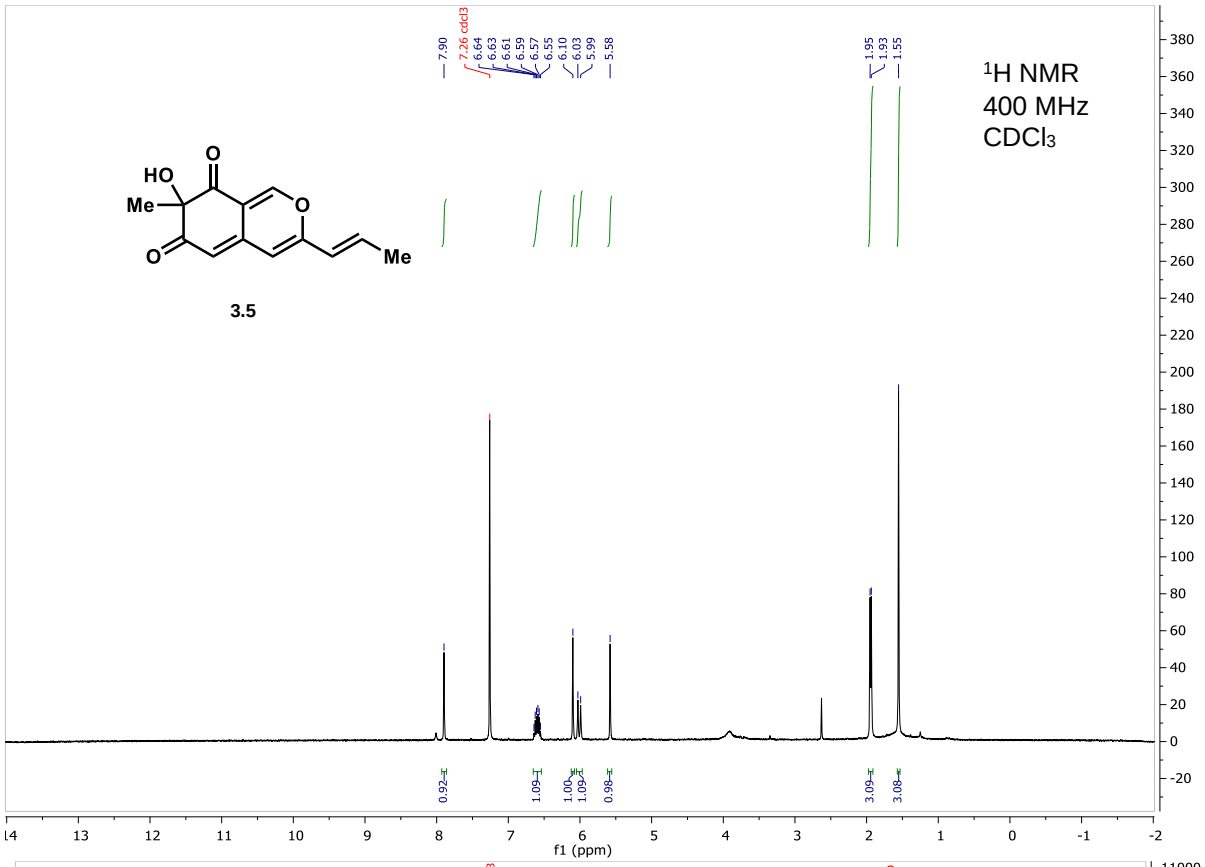

160.98
159.24
158.90
152.21
139.14
131.57
122.93
122.33
107.36
105.06
103.70
101.62
94.16
77.22 cdcl3
76.80 cdcl3
57.74
56.40
18.33
9.88
13C NMR
600 MHz
CDCl3
MOMO
O
Me
O
MOMO
Me
3.S7
f1 (ppm)

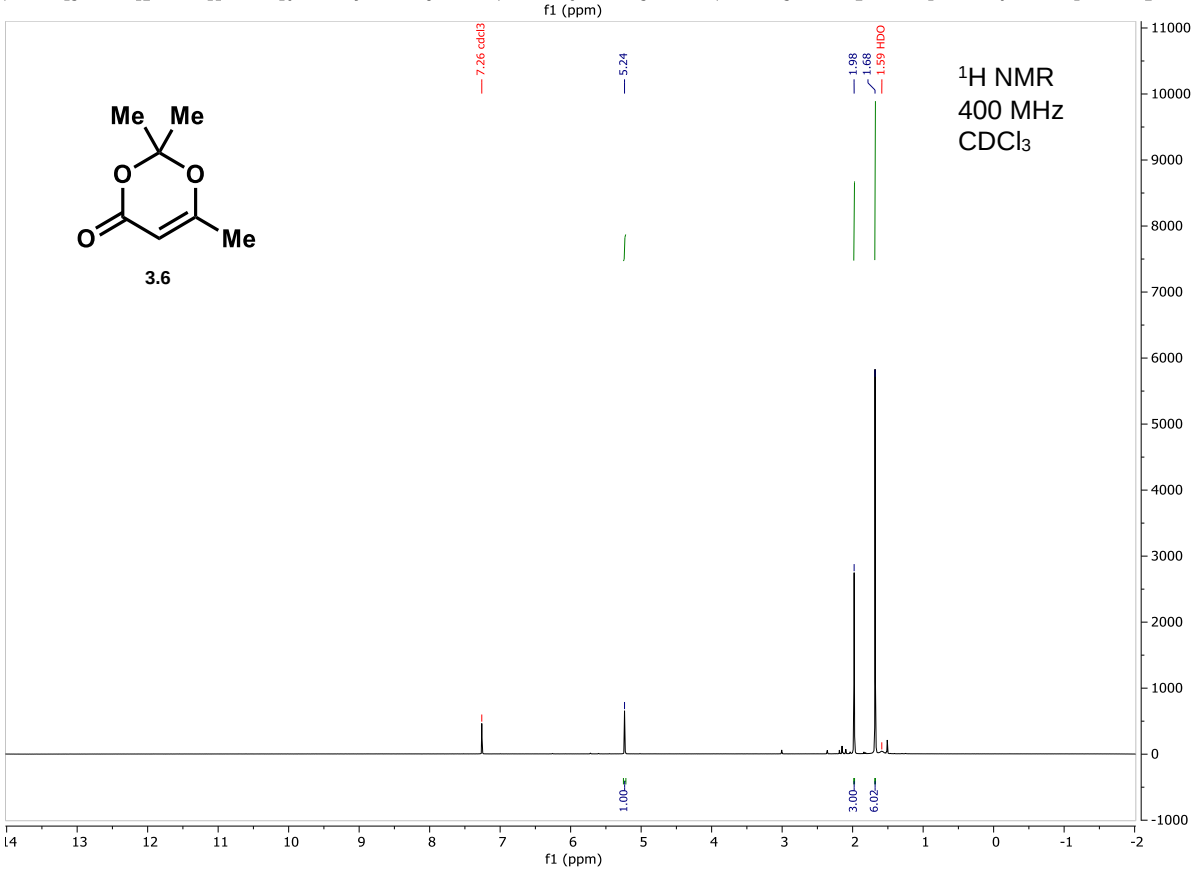

6.63
6.57
6.56
6.34
6.32
6.30
5.99
5.99
5.95
5.95
5.78
5.16
5.01
3.60
3.45
2.15
1.84
1.82 HDO
1.82
1H NMR
400 MHz
CDCl3
MOMO
OH
Me
O
MOMO
Me
3.S8
1.02
1.10
1.01
1.13
1.00
2.01
1.97
2.78
2.95
3.00
3.01
f1 (ppm)

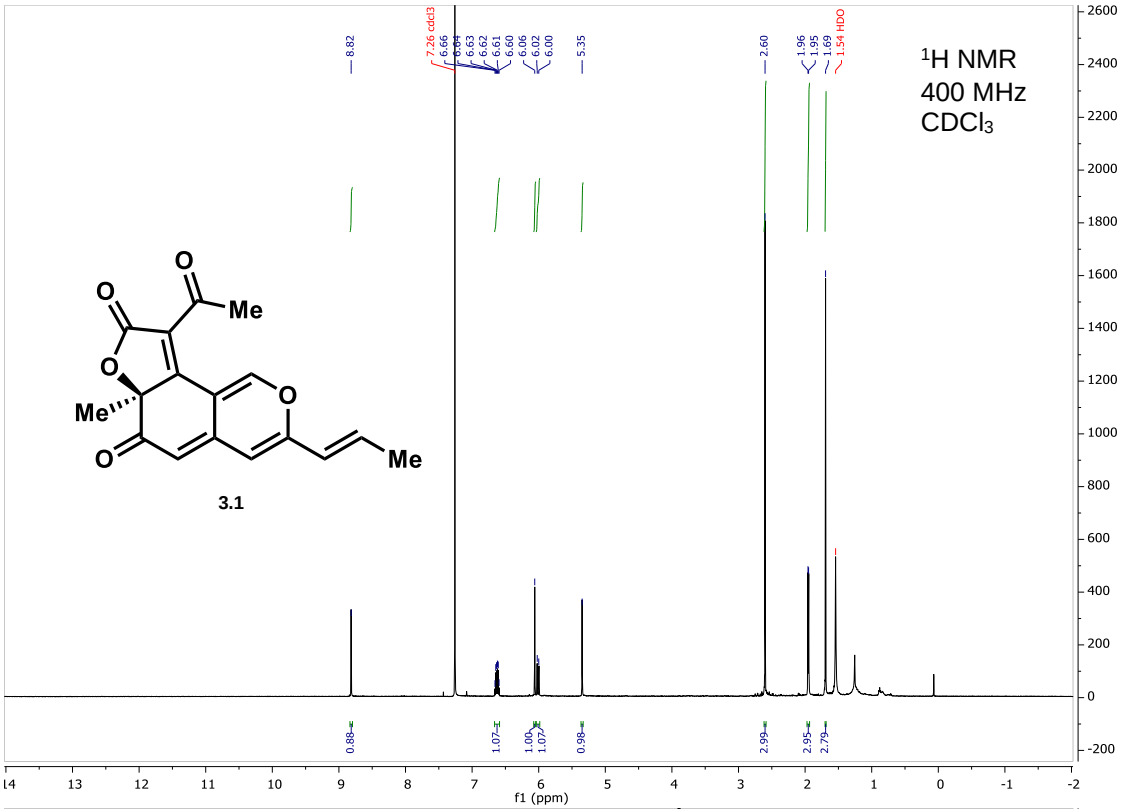

156.69
152.96
148.72
129.12
128.28
126.00
118.98
116.41
105.79
102.03
99.91
94.39
88.95
77.45 cdcl3
77.13 cdcl3
76.81 cdcl3
57.54
56.00
18.16
9.90
13C NMR
600 MHz
CDCl3
MOMO
OH
Me
O
MOMO
Me
3.S8
f1 (ppm)

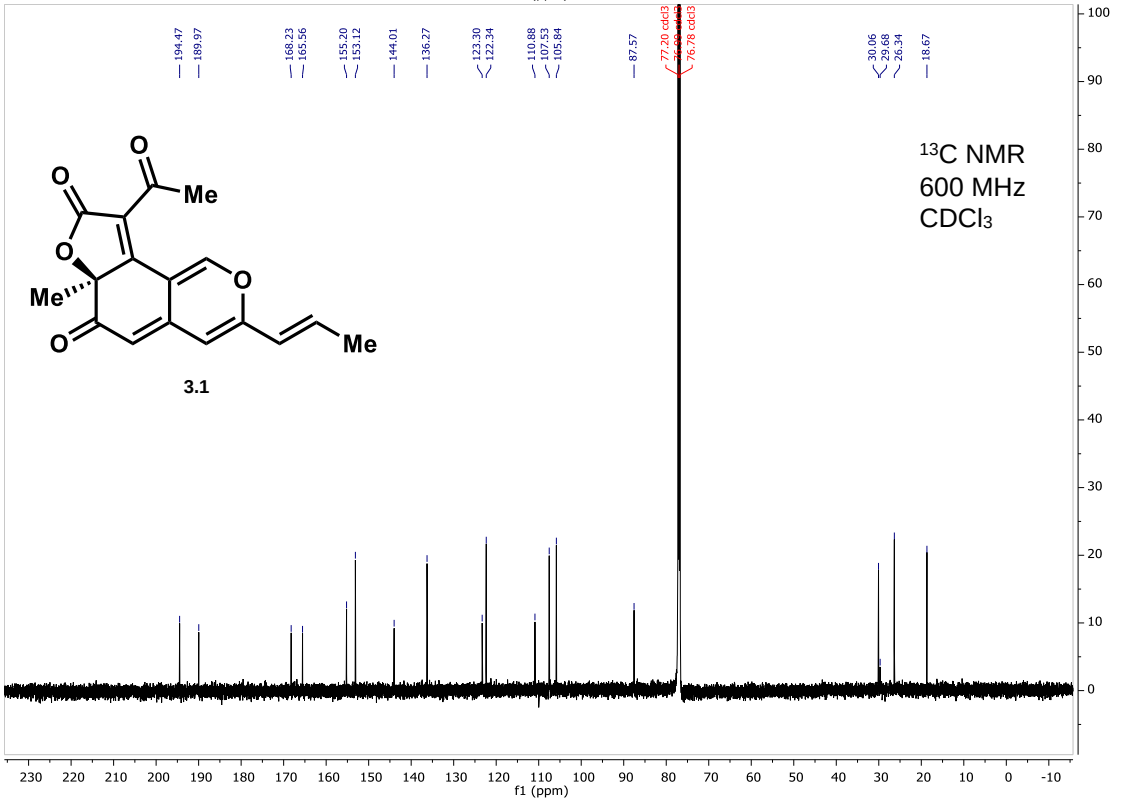

HO
O
Me
O
HO
Me
3.4
^{1}H NMR
400 MHz
CD_3OD
f1 (ppm)

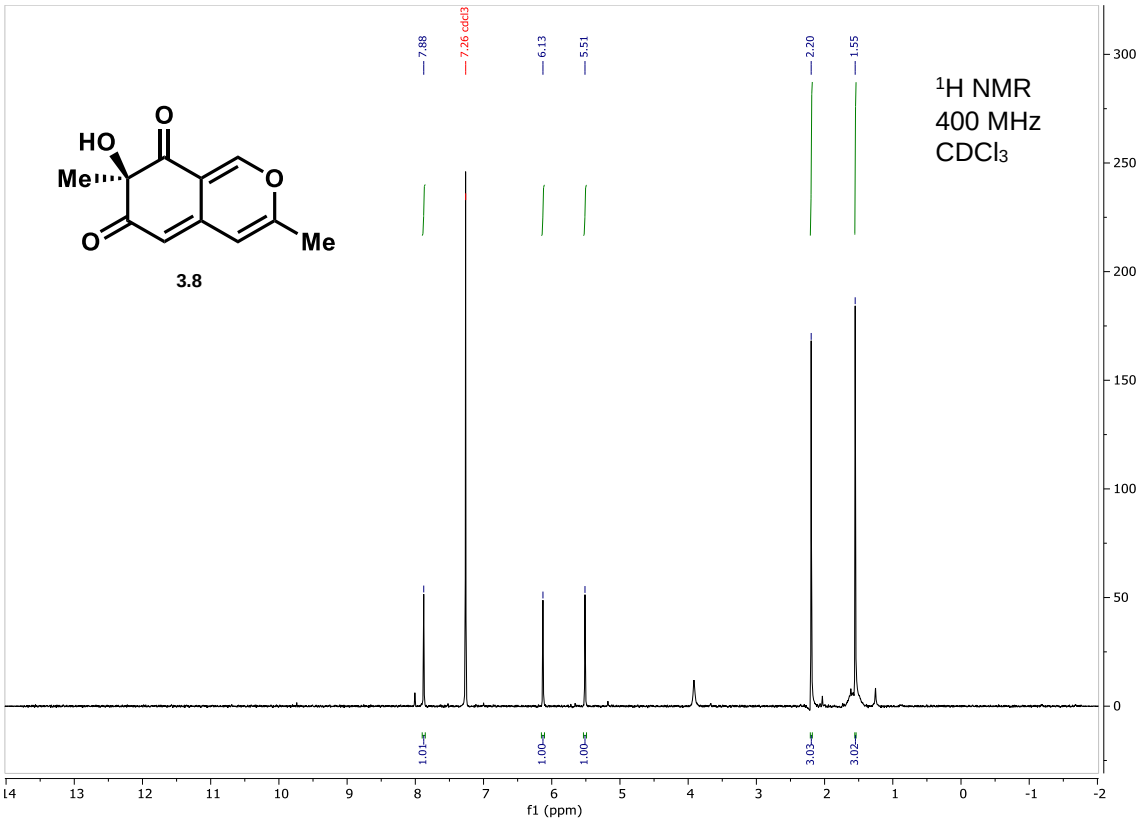

HO
O
Me
O
HO
Me
3.4
^{13}C NMR
600 MHz
CD_3OD
f1 (ppm)

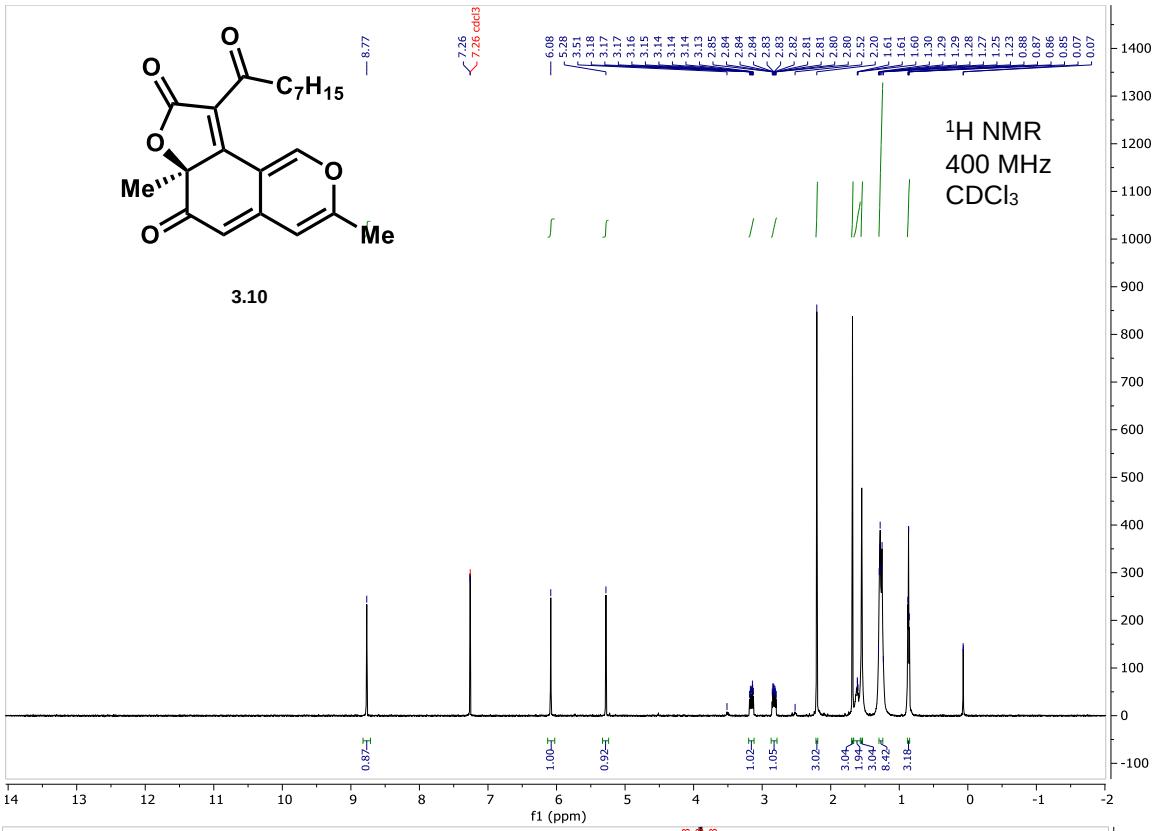

HO
Me
O
O
O
Me
3.5
7.90
7.26 cdcl3
6.64
6.63
6.61
6.59
6.57
6.55
6.10
6.03
5.99
5.58
1.95
1.93
1.55
0.92
1.09
1.00
1.09
0.98
3.09
3.08
^{1}H NMR
400 MHz
$CDCl_3$
f1 (ppm)

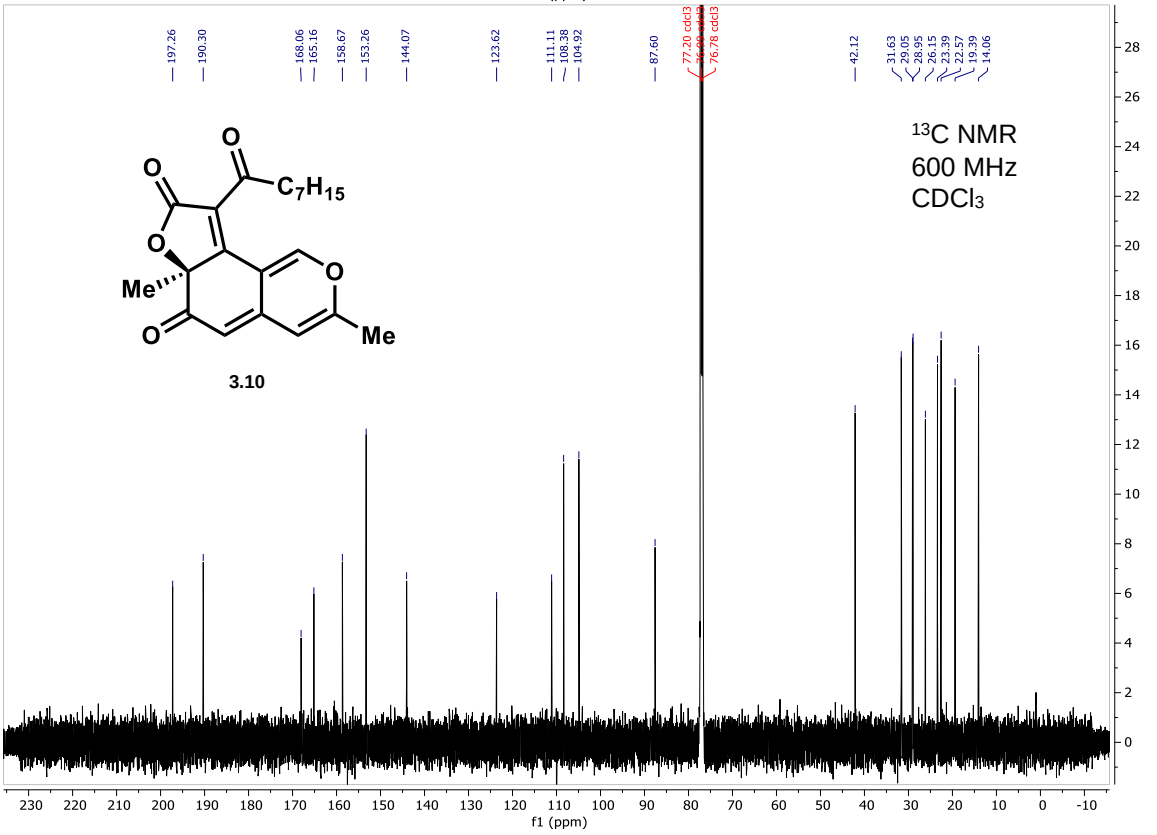

Me
Me
O
O
O
Me
3.6
7.26 cdcl3
5.24
1.98
1.68
1.59 HDO
1.00
3.00
6.02
^{1}H NMR
400 MHz
$CDCl_3$
f1 (ppm)

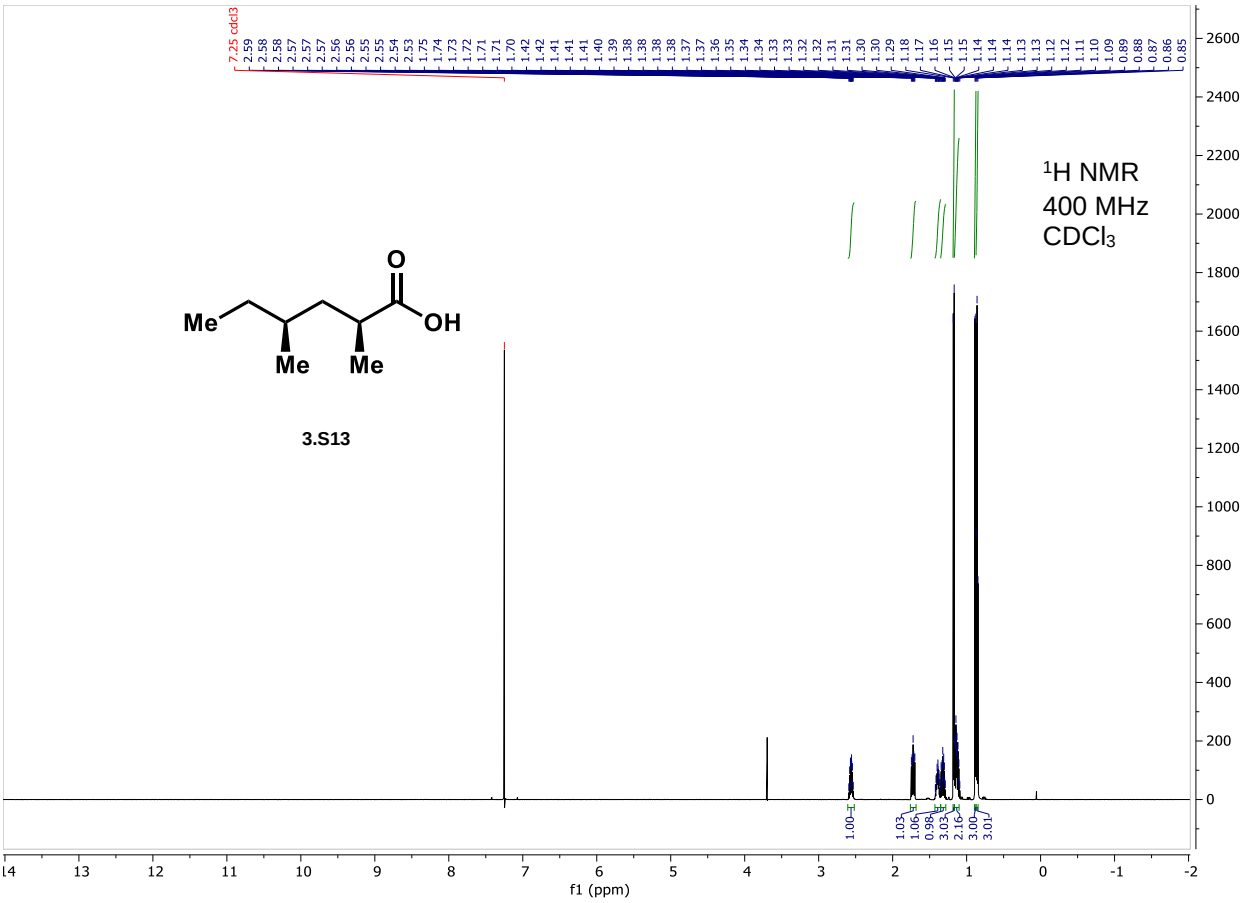

^{1}H NMR
400 MHz
$CDCl_3$
3.1
f1 (ppm)

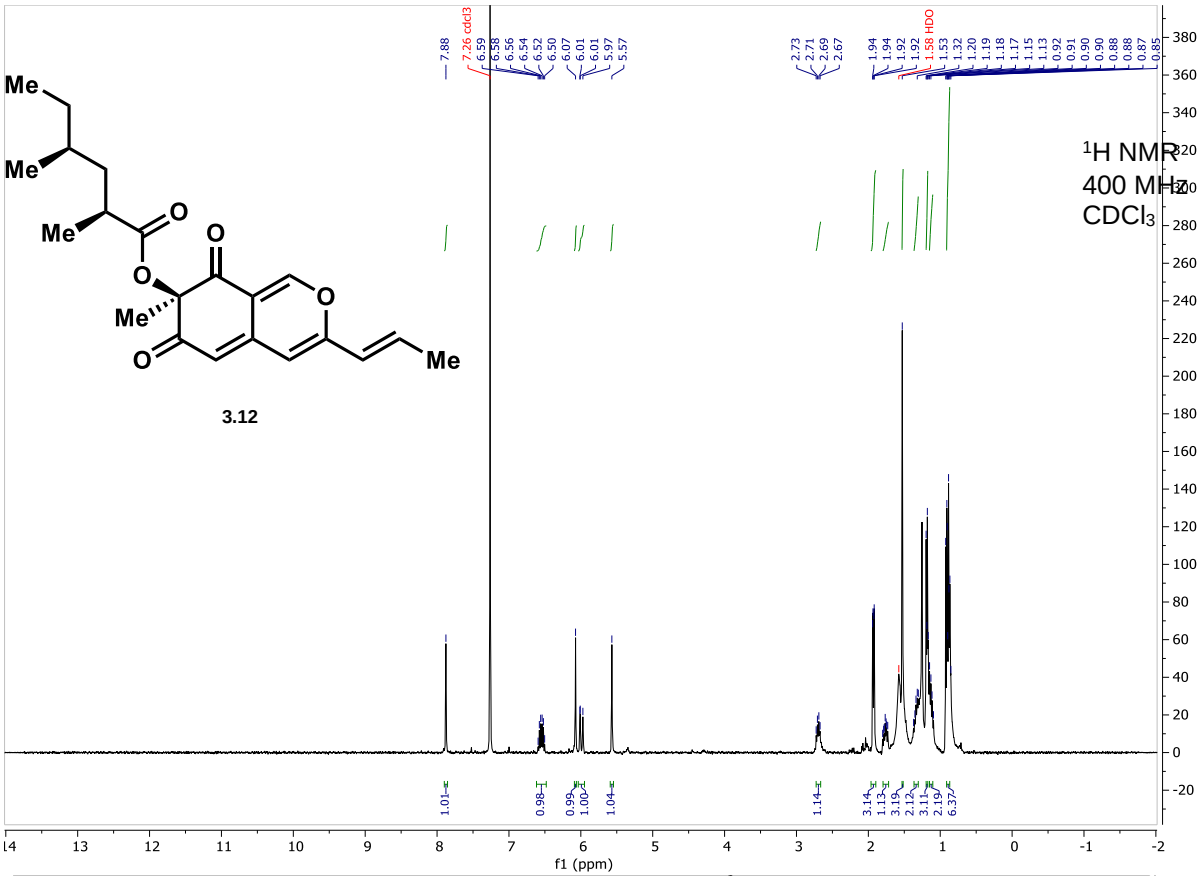

^{13}C NMR
600 MHz
$CDCl_3$
3.1
f1 (ppm)

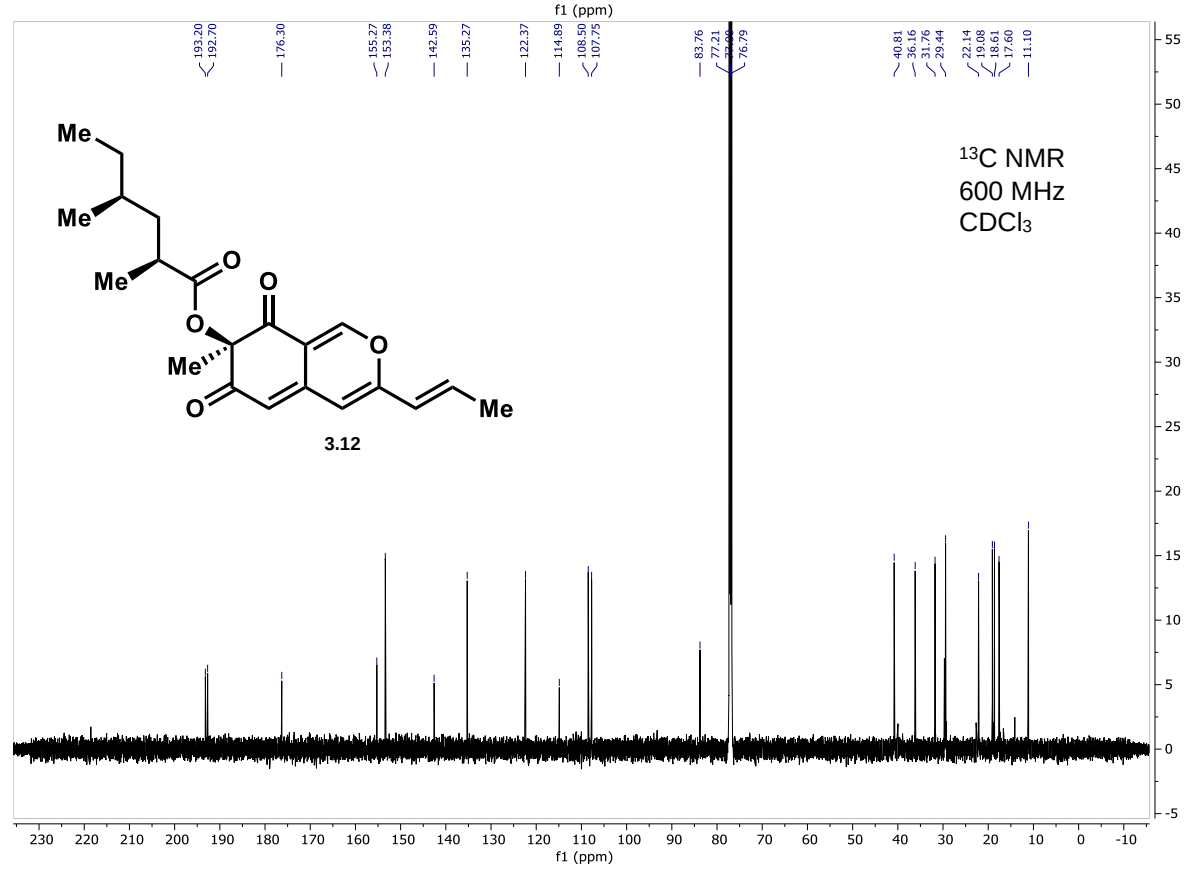

HO
Me
O
O
O
Me
3.8
^{1}H NMR
400 MHz
CDCl$_3$
7.88
7.26 cdcl3
6.13
5.51
2.20
1.55
1.01
1.00
1.00
3.03
3.02
f1 (ppm)

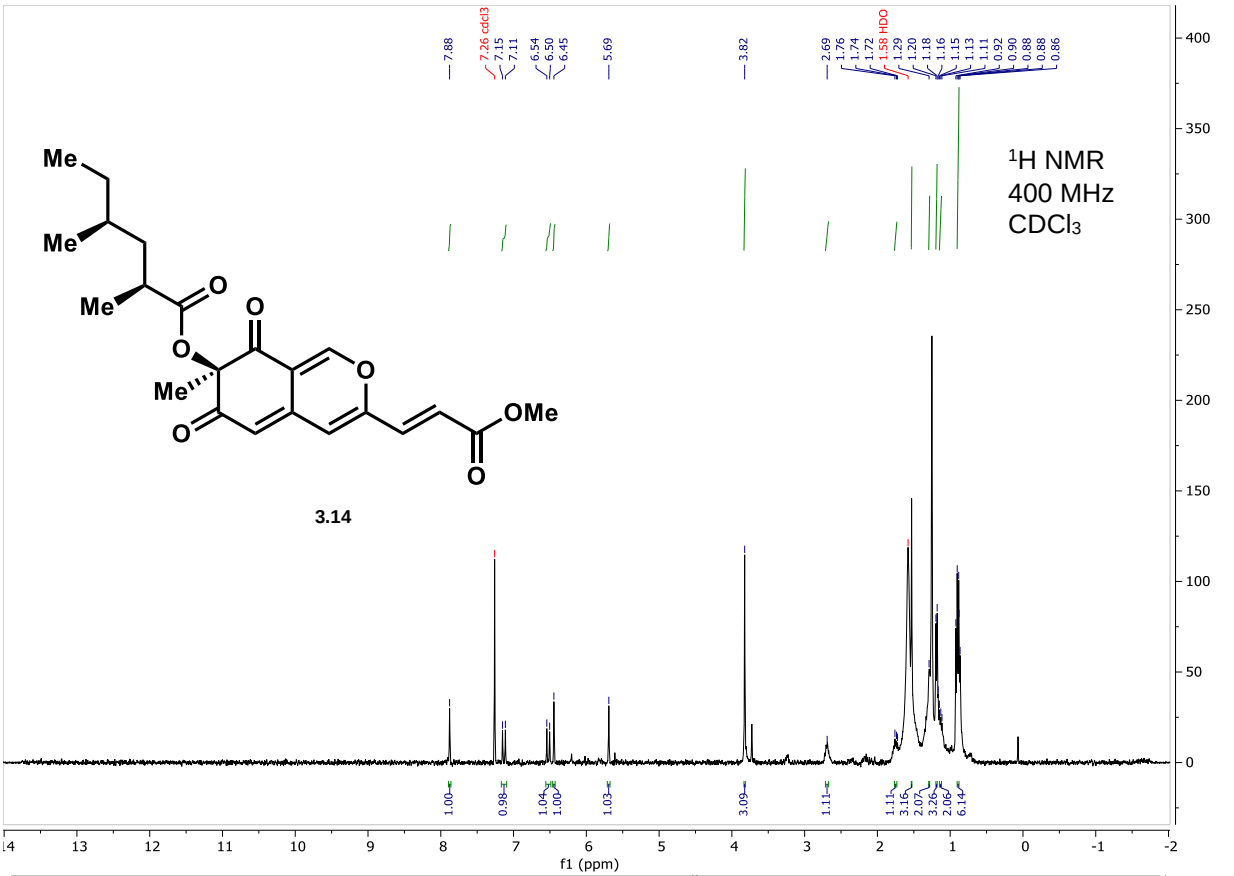

^{1}H NMR
400 MHz
$CDCl_3$
3.10
f1 (ppm)

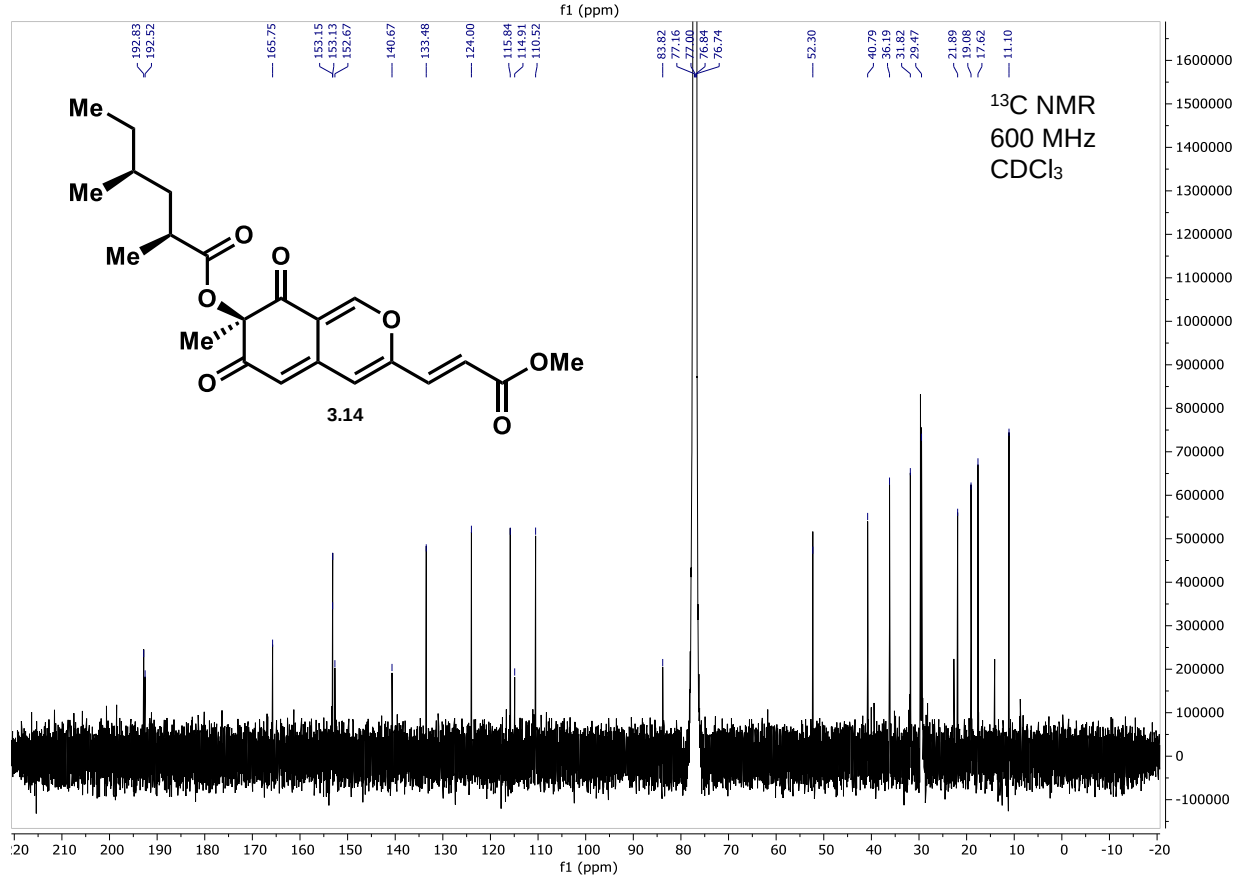

^{13}C NMR
600 MHz
$CDCl_3$
3.10
f1 (ppm)


^{1}H NMR
400 MHz
$CDCl_3$
Me
Me
Me
O
OH
3.S13
f1 (ppm)


Me
Me
Me
O
O
O
Me
O
O
Me
3.12
^{1}H NMR
400 MHz
$CDCl_3$
f1 (ppm)


Me
Me
Me
O
O
O
Me
O
O
Me
3.12
^{13}C NMR
600 MHz
$CDCl_3$
f1 (ppm)


7.88
7.26 cdcl3
7.15
7.11
6.54
6.50
6.45
5.69
3.82
2.69
1.76
1.74
1.72
1.58 HDO
1.29
1.20
1.18
1.16
1.15
1.13
1.11
0.92
0.90
0.88
0.88
0.86
1H NMR
400 MHz
CDCl3
Me
Me
Me
O
O
O
Me
O
O
OMe
O
3.14
1.00
0.98
1.04
1.00
1.03
3.09
1.11
1.11
3.16
2.07
3.26
2.06
6.14
f1 (ppm)

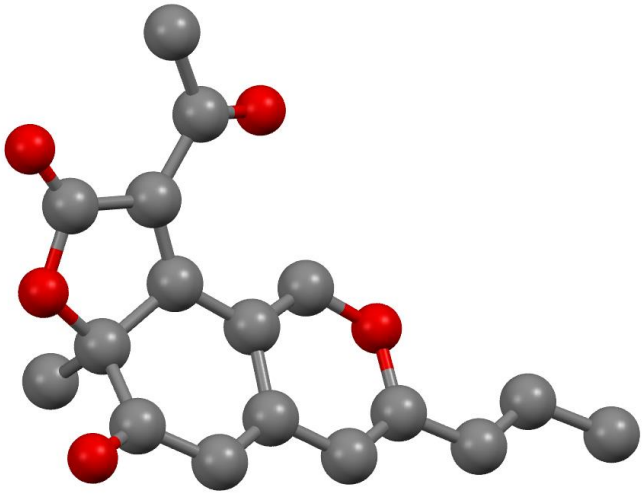

192.83
192.52
165.75
153.15
153.13
152.67
140.67
133.48
124.00
115.84
114.91
110.52
83.82
77.16
77.00
76.84
76.74
52.30
40.79
36.19
31.82
29.47
21.89
19.08
17.62
11.10
13C NMR
600 MHz
CDCl3
Me
Me
Me
O
O
O
Me
O
O
OMe
O
3.14
f1 (ppm)

VII. Single-Crystal Structure Determination

Single-crystal X-ray diffraction data were collected using a Rigaku XtaLAB Synergy-S X-ray diffractometer configured in a kappa goniometer geometry. The diffractometer is equipped with a PhotonJet-S microfocus Cu source (λ = 1.54187 Å) set at a rough divergence of 9.5 and operated at 50 kV and 1 mA. X-ray intensities were measured at 298(1) K with the HyPix-6000HE detector placed 34.00 mm from the sample. The data were processed with CrysAlisPro v38.46 (Rigaku Oxford Diffraction) and corrected for absorption. The structures were solved in OLEX2[26] using SHELXTL[27] and refined using SHELXL.[28] All non-hydrogen atoms were refined anisotropically with hydrogen atoms placed at idealized positions.

Table 3.2 Crystallographic parameters for trichoflectin

Material	exp_589
Space Group	$P2_12_12_1$
a Å	5.39440(6)
b Å	18.0328(2)
c Å	29.9372(4)
Volume (Å^3)	2912.18
Flack	0.00(6)
Temperature	298(1)
ρ_{calc} (g cm^{-3})	1.361
R_1/wR_2	3.00/8.46
GOF	1.052



ORTEP Diagram

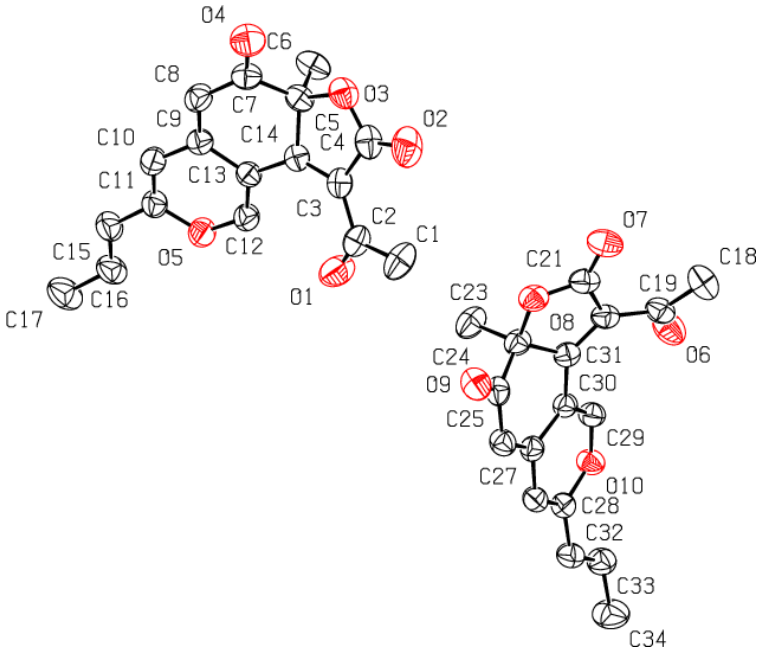


Figure 3.12. A view of exp_589 showing the atom labelling scheme. Thermal ellipsoids are drawn at the 50% probability level and H atoms are omitted for clarity.

VIII. Circular Dichroism Spectroscopy of Azaphilones 3.1, 3.10, 3.14

Instruments and methods

Circular dichroism (CD) spectra were collected on a J-1500 Circular Dichroism Spectrophotometer (Jasco). Samples were prepared in UPLC grade MeCN at a final concentration of 750 μM. Data points were collected at room temperature from 190 to 500 nm with a scan rate of 100 nm/min in a quartz cell with an optical path of 1 mm.

ECD Methodology

The general approach for absolute configuration assignment using ECD, including the detailed computational workflow, has been published elsewhere.[29-31] A subset of the details of the computational methodology is provided here. Conformers of each test structure were geometry optimized at the B3LYP/6-31G** level and stationary points were confirmed by performing frequency calculations.[32-40] All calculations were performed using Gaussian 09.[41] Output conformers were ranked according to DFT energy and a clustering was performed in order to remove duplicates. Initial duplicate identification was performed solely on an electronic energy basis where two compounds were considered identical if the difference in Hartrees was less than 0.01. Rounding the differences led to inconsistencies in identification of duplicates. It became better to cluster the DFT minima by energy and then re-cluster each energy bucket by structure using an all atom RMS of 0.6 Å. This process removed just identical compounds. Two Boltzmann distributions were calculated based on the free energy (G) and the electronic energy (E).
To calculate UV and ECD spectra, B3LYP geometries were used as input. The spectra were then calculated using either the B3LYP or CAM-B3LYP[42] functionals, along with the 6-31++G** basis set[43, 44] in vacuo. Only conformers which contributed more than 5.0% to the total in vacuo conformer distribution were selected for UV and ECD calculation. Time-dependent Density Functional Theory (TDDFT)[45] methodology was employed using the following keywords: TD=full,singlet, Nstates=100, and integral=ultrafinegrid. Spectral display, Boltzmann weighting, and curve fitting were carried out using SpecDis,[46, 47] and were displayed with a wavelength shift and band broadening sigma values in order to best match the calculated and experimental UV spectra. This shift and band broadening were then applied to the ECD spectra, and the area under the curve fit was determined by SpecDis.

Trichoflectin (**3.1**)
Calculations of the ECD and UV spectra (CAM-B3LYP/6-31++G**) involved modeling the (*R*)- enantiomer of the natural product. Since no other stereoisomers were possible, it should be noted that the (*S*)-enantiomer is assumed to have a spectrum that will be equal and opposite at all wavelengths.
Figure S22 provides an overlay of the calculated and measured UV and ECD spectra using the theoretical spectrum of the (*R*)-enantiomer. The calculated spectrum has been shifted 30 nm and a band broadening of σ=0.39 eV applied in order to optimize the UV spectral match. A high degree of confidence is derived from the statistical and visual matching of the spectra, with the experimental spectrum matching the mirror image of the

calculated (*R*)-enantiomer. Hence, the absolute configuration of trichoflectin can confidently be assigned as (*S*). Figure S23 shows the one conformer of the *(R)* enantiomer that contributed >5% to the Boltzmann weighted spectrum, and the coordinates of this conformer are shown with the electronic energy below. Assigned absolute configuration of the desired compounds is provided in Scheme S1.

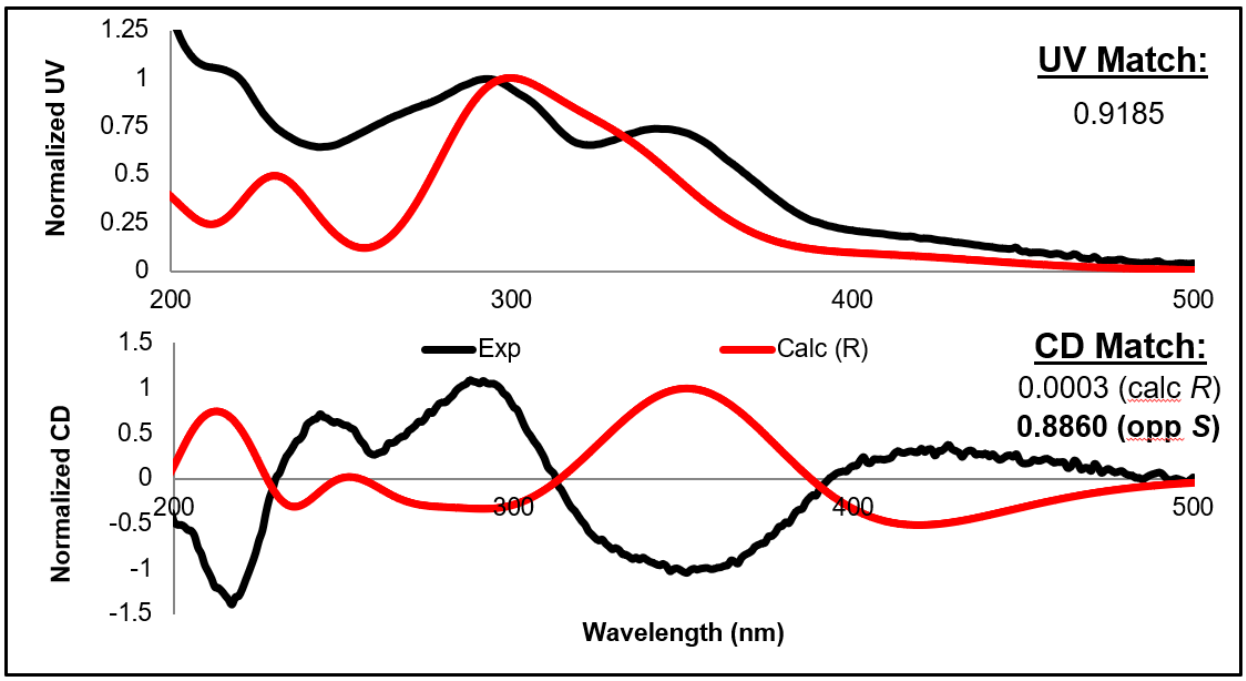


Figure 3.13. Comparison between experimental (black) and calculated (red) UV (top) and ECD (bottom) spectra. The calculated (R)-enantiomer is opposite of the experimental spectrum, with a large difference in fits (Δ=0.8857) suggesting a confident assignment. The calculated spectrum has been shifted 30 nm, and a band broadening of 0.39 eV has been applied

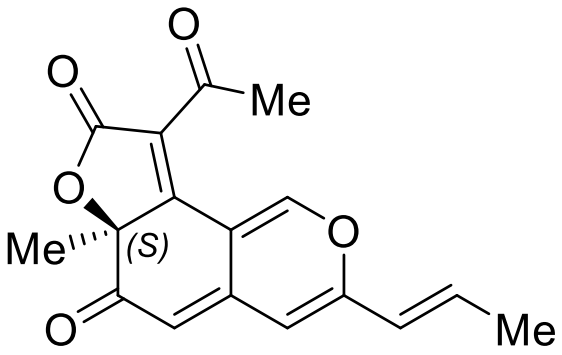
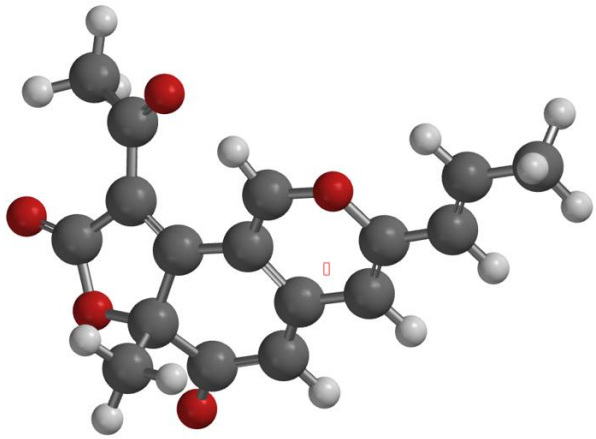


Figure 3.14. Assigned absolute configuration of trichoflectin based on ECD analysis



Conformer 1: 93.3%

Figure 3.15. One conformer of the (*R*)-enantiomer of trichoflectin that contributes >5% to the Boltzmann distribution. Note the percentage shown above based is on *in vacuo* electronic energies.

Coordinates and electronic energies for B3LYP/6-31G conformational minima contributing >5% to the *in vacuo* Boltzmann distribution.**

Conformer 1: -1032.346471 hartrees

6	-1.92652	-1.24179	0.454749	6	-2.84313	3.30205	-0.90379
6	-1.43886	-2.36436	-0.50656	6	-1.81538	-1.7794	1.89715
6	-0.00398	-2.37628	-0.72861	1	7.0976	1.54991	-0.61613
6	0.811717	-1.3366	-0.3621	1	7.1483	-0.21358	-0.41144
6	0.287916	-0.07181	0.177876	1	7.22791	0.866265	0.996514
6	-1.16113	0.044932	0.262781	1	-2.48918	4.32822	-1.00616
6	1.15752	0.916672	0.51399	1	-3.79449	3.26342	-0.3666
6	2.24341	-1.37366	-0.51756	1	-3.0406	2.8698	-1.89086
8	-2.21691	-3.20925	-0.91931	1	-2.17304	-1.0227	2.60047
6	3.04043	-0.3412	-0.14027	1	-0.78389	-2.04131	2.14313
6	4.48685	-0.31902	-0.24494	1	-2.44014	-2.67163	1.98075
8	2.49748	0.804537	0.396205	1	0.408491	-3.25877	-1.20725
6	5.27154	0.705286	0.129229	1	0.849272	1.88074	0.892321
6	-2.0555	1.05447	0.067487	1	2.70889	-2.25918	-0.93574
6	-3.41789	0.444905	0.026975	1	4.92704	-1.22041	-0.66473
8	-3.29409	-0.90399	0.221383	1	4.80462	1.59524	0.545569
6	-1.79661	2.49736	-0.17149				
8	-0.75437	3.01999	0.214502				
8	-4.4906	0.97176	-0.14711				
6	6.76245	0.716682	0.01421				

Lunatoic Acid A Methyl Ester (**3.14**)

Calculations of the ECD and UV spectra (CAM-B3LYP/6-31++G**) involved modeling the (*R*) enantiomer of the natural product methyl ester. While the compound contains an ester chain with two stereocenters, the majority of the UV absorption and thus the ECD signal is expected to come from the stereocenter on the ring. Calculations of a hypothetical diastereomer with inversion at the stereocenter alpha to the ester carbonyl showed an analogous signal to that for the expected structure, indicating that this prior assumption is correct.

Figure S26 provides overlays of the calculated and measured UV and ECD spectra using the theoretical spectrum of the (*R*)-enantiomer. The calculated spectrum has been shifted -46 nm and a band broadening of σ=0.28 eV applied in order to optimize the UV spectral match. A high degree of confidence is derived from the statistical and visual matching of the spectra, with the experimental spectrum matching the calculated (*R*)-enantiomer. Hence, the absolute configuration of the methyl ester of Lunatoic Acid A can confidently be assigned as (*R*). Figure S27 shows the six conformers of the *(R)* enantiomer that contributed >5% to the Boltzmann weighted spectrum, and the coordinates of these conformers are shown with the electronic energies listed below. Assigned absolute configuration of the desired compounds is provided in Scheme S2.

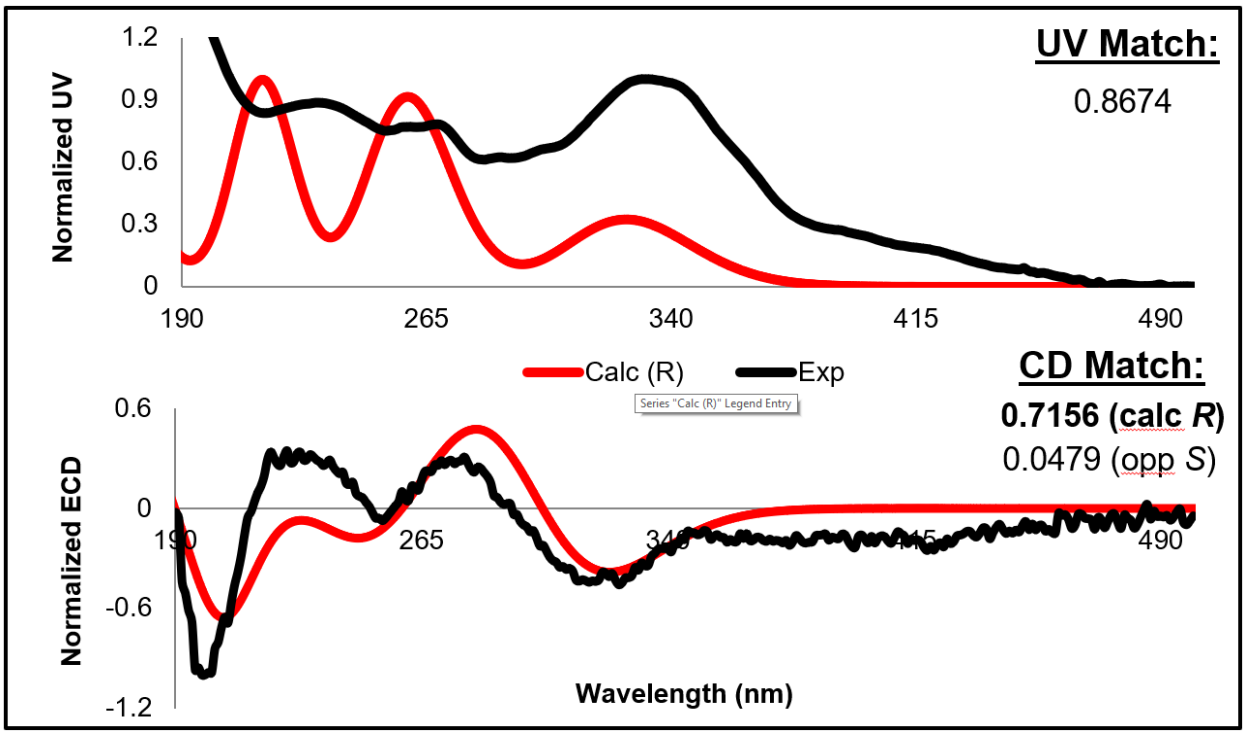


Figure 3.16. Comparison between experimental (black) and calculated (red) UV (top) and ECD (bottom) spectra. The calculated (*R*)-enantiomer is a good match to the experimental, with a large difference in fits (Δ=0.6677) suggesting a confident assignment. The calculated spectrum has been shifted -46 nm, and a band broadening of 0.28 eV has been applied.

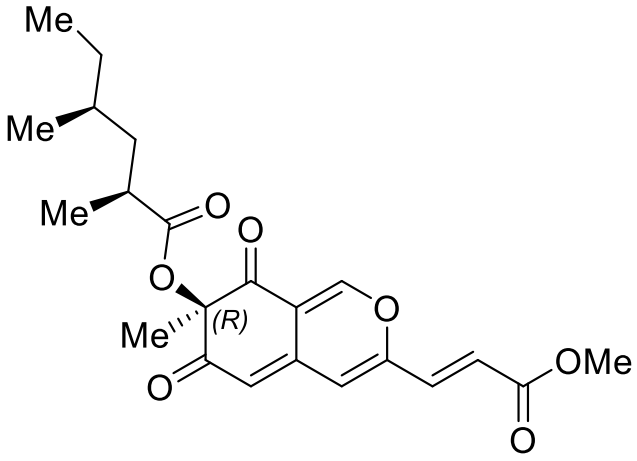


Figure 3.17. Assigned absolute configuration of lunatoic acid A methyl ester based on ECD analysis.

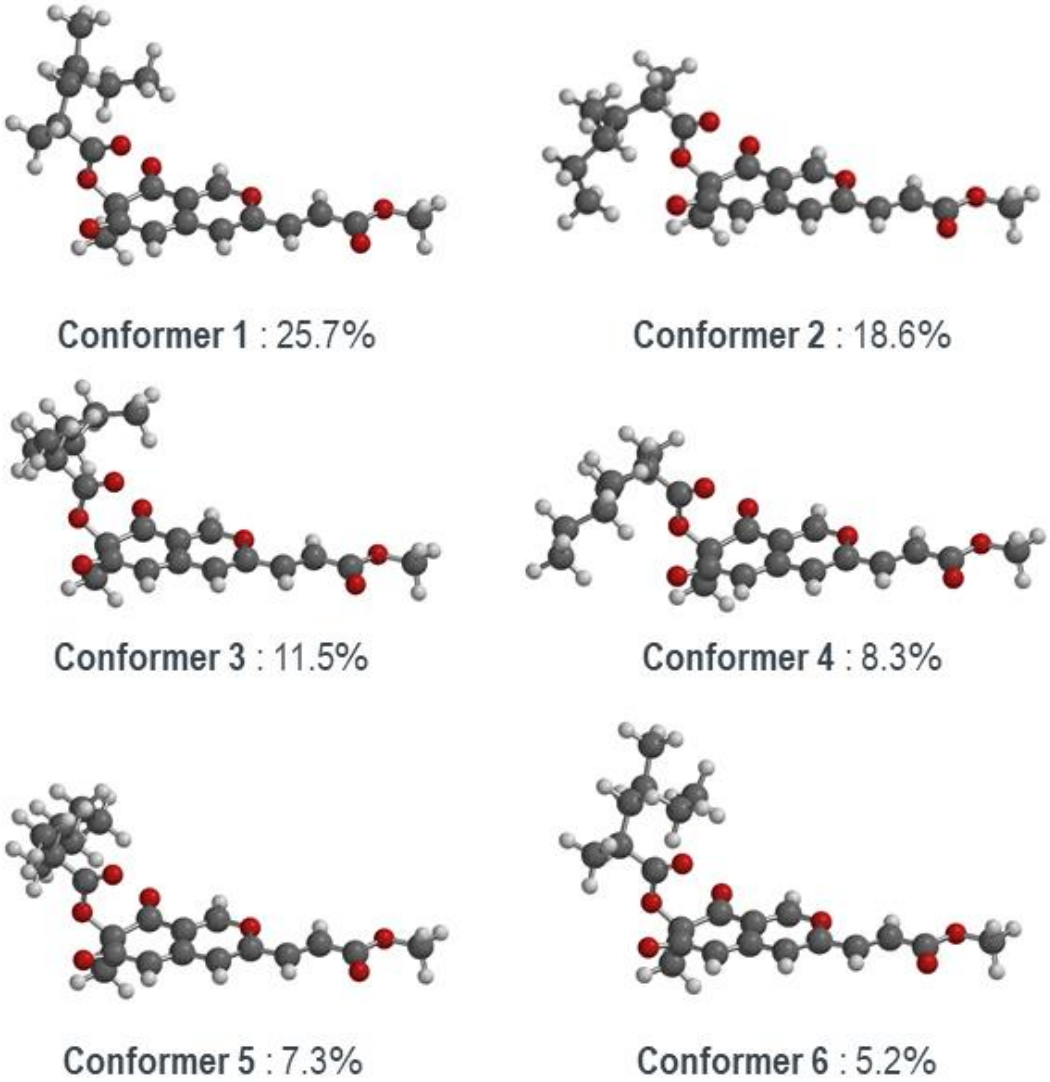


Figure 3.18. Six conformers of the (*R*)-enantiomer of lunatoic acid A methyl ester that contribute >5% to the Boltzmann distribution. Note the percentage shown above based is on *in vacuo* electronic energies.

Coordinates and electronic energies for B3LYP/6-31G conformational minima contributing >5% to the *in vacuo* Boltzmann distribution.**

Conformer 1: -1380.662338 hartrees

6	1.78282	-1.68543	0.941074
6	1.32039	-2.58496	-0.23991
6	-0.08437	-2.56709	-0.60377
6	-0.9426	-1.60841	-0.14681
6	-0.5196	-0.5554	0.767273
6	0.874425	-0.48446	1.26383
6	-2.32929	-1.52556	-0.54115
6	-3.1488	-0.54276	-0.08739
8	-2.69057	0.431987	0.778282
6	-1.39911	0.390925	1.16506
6	-4.54266	-0.40614	-0.4433
6	-5.36122	0.565606	-0.00381
6	-6.77309	0.588333	-0.44821
8	-7.42645	1.63913	0.094655
8	-7.29006	-0.21849	-1.19745
6	-8.81115	1.75329	-0.27781
8	1.27673	0.428157	1.96608
8	2.14123	-3.32308	-0.7703
6	1.88912	-2.55832	2.19714
8	3.11592	-1.21925	0.681905
6	3.21769	-0.42005	-0.4067
6	4.65684	-0.13761	-0.77387
8	2.24322	-0.05119	-1.03495
6	4.86859	1.32854	-1.19008
6	5.00568	2.33153	-0.02596
6	3.77038	2.4135	0.898726
6	5.40257	3.71242	-0.57045
6	5.00231	-1.11961	-1.91571
6	2.48952	2.96521	0.26043
1	-9.18284	2.6388	0.236805
1	-9.36624	0.864184	0.032778
1	-8.90866	1.86616	-1.36077
1	2.26945	-1.95226	3.02238
1	2.57353	-3.38374	1.99027
1	0.910116	-2.96483	2.46741
1	5.50425	4.44226	0.240151
1	4.65198	4.09497	-1.27036
1	6.35778	3.66779	-1.10507
1	6.06683	-1.05241	-2.15839
1	4.42381	-0.8727	-2.81099
1	4.76711	-2.14983	-1.63287
1	1.67012	2.92449	0.983973
1	2.60967	4.00572	-0.05882
1	2.18371	2.37216	-0.60516
1	4.05069	1.6223	-1.85843
1	5.78692	1.37485	-1.78889
1	4.0386	3.04453	1.75683
1	3.55559	1.42696	1.32415
1	5.27726	-0.37544	0.096936
1	5.83819	1.97483	0.600279
1	-0.4089	-3.31549	-1.31925
1	-2.73808	-2.26244	-1.22363
1	-1.13259	1.20076	1.8328
1	-4.9434	-1.15583	-1.12031
1	-5.02585	1.34477	0.670543

Conformer 2: -1380.662034 hartrees

6	1.60301	-0.03006	1.10298
6	1.3789	-1.37343	0.355659
6	0.014003	-1.7783	0.076571
6	-1.04379	-0.9249	0.210908
6	-0.88031	0.452365	0.656243
6	0.458903	0.995158	0.987229
6	-2.40696	-1.28817	-0.0968
6	-3.43234	-0.40746	0.032594
8	-3.21705	0.887649	0.463061
6	-1.95435	1.26785	0.747312
6	-4.81605	-0.71065	-0.25362
6	-5.83805	0.154289	-0.13215
6	-7.21264	-0.28798	-0.45871
8	-8.09295	0.721015	-0.27846
8	-7.52427	-1.40097	-0.83766
6	-9.4627	0.393566	-0.57157
8	0.650022	2.17255	1.23702
8	2.3459	-2.0847	0.107215
6	1.81116	-0.3468	2.59071

8	2.82306	0.578009	0.669392
6	2.82824	0.976139	-0.62908
6	4.13858	1.63788	-1.00285
8	1.86405	0.836693	-1.35396
6	5.34521	0.822473	-0.50325
6	5.39588	-0.62932	-1.02382
6	6.2899	-1.51713	-0.1358
6	5.85781	-0.67645	-2.48719
6	4.13868	3.07093	-0.43476
6	5.70701	-1.793	1.25639
1	-10.0311	1.3026	-0.37759
1	-9.80742	-0.42213	0.069435
1	-9.56843	0.086968	-1.61553
1	2.01622	0.581948	3.12786
1	2.65955	-1.02784	2.68593
1	0.921112	-0.82186	3.01336
1	5.82731	-1.6996	-2.87624
1	6.88894	-0.31306	-2.58176
1	5.22695	-0.06024	-3.13642
1	5.05623	3.58945	-0.72908
1	4.0866	3.04841	0.657594
1	3.2833	3.64439	-0.80353
1	6.33759	-2.49318	1.81458
1	5.62542	-0.87924	1.85454
1	4.70273	-2.219	1.16926
1	5.31673	0.82973	0.590032
1	6.26441	1.34783	-0.79631
1	7.28637	-1.05919	-0.04828
1	6.44131	-2.47593	-0.64853
1	4.13371	1.68799	-2.09586
1	4.38433	-1.05266	-0.96904
1	-0.12316	-2.78673	-0.29964
1	-2.63015	-2.2909	-0.44405
1	-1.88497	2.30284	1.05848
1	-5.02839	-1.72114	-0.59244
1	-5.69587	1.17646	0.198431

Conformer 3: -1380.661580 hartrees

6	1.80645	-1.44169	0.96146
6	1.25025	-2.64964	0.155399
6	-0.17932	-2.70503	-0.09509
6	-0.9995	-1.63538	0.118238
6	-0.5044	-0.36532	0.63408
6	0.927277	-0.17683	0.961251
6	-2.41536	-1.64472	-0.16665
6	-3.19794	-0.55479	0.039268
8	-2.67158	0.624928	0.532123
6	-1.35133	0.67509	0.80152
6	-4.61807	-0.50162	-0.22184
6	-5.40103	0.573276	-0.02404
6	-6.84641	0.491423	-0.33301
8	-7.45656	1.66853	-0.07229
8	-7.42178	-0.4906	-0.76212
6	-8.86935	1.69464	-0.34157
8	1.38571	0.902317	1.30141
8	2.02181	-3.54237	-0.16765
6	2.00108	-1.88773	2.41606
8	3.1193	-1.11135	0.482608
6	3.14307	-0.63909	-0.78697
6	4.53741	-0.2664	-1.24181
8	2.13028	-0.48912	-1.44328
6	4.66904	1.27368	-1.1446
6	4.59209	1.85383	0.282993
6	4.31946	3.37107	0.257395
6	5.87038	1.54384	1.07516
6	4.76159	-0.76217	-2.67659
6	2.91313	3.7473	-0.22612
1	-9.20045	2.69967	-0.08196
1	-9.389	0.947829	0.264516
1	-9.06233	1.48472	-1.39691
1	2.44538	-1.06879	2.98614
1	2.66392	-2.75546	2.42639
1	1.04297	-2.16464	2.86562
1	5.79709	1.92248	2.1002
1	6.74457	2.01411	0.607319
1	6.06246	0.46803	1.13768
1	5.75721	-0.4714	-3.02445
1	4.01456	-0.32844	-3.34689
1	4.67943	-1.85142	-2.73795
1	2.74341	4.82575	-0.13886
1	2.75264	3.47711	-1.27494
1	2.15184	3.22853	0.365762
1	3.87972	1.70055	-1.77273
1	5.62504	1.56475	-1.60085
1	5.07915	3.86872	-0.36329

1	4.4569	3.76179	1.27411
1	5.25224	-0.74403	-0.56586
1	3.74888	1.38755	0.806291
1	-0.55972	-3.62808	-0.52008
1	-2.87746	-2.54332	-0.56017
1	-1.03229	1.64273	1.16855
1	-5.07182	-1.41047	-0.60798
1	-5.01289	1.51069	0.356631

Conformer 4: -1380.661272 hartrees

6	-1.55596	0.525882	-1.24688
6	-1.39326	-1.0005	-1.00539
6	-0.05116	-1.52379	-0.8291
6	1.02845	-0.71386	-0.622
6	0.913797	0.737612	-0.57223
6	-0.39731	1.40781	-0.74319
6	2.36829	-1.20795	-0.4078
6	3.41619	-0.37277	-0.18895
8	3.24754	0.998342	-0.15597
6	2.00729	1.49677	-0.33804
6	4.77861	-0.80378	0.026419
6	5.82249	0.014168	0.247073
6	7.1711	-0.5609	0.452202
8	8.07907	0.417463	0.662086
8	7.44234	-1.74653	0.437014
6	9.42675	-0.03819	0.875153
8	-0.55242	2.60508	-0.57627
8	-2.38406	-1.71879	-1.06503
6	-1.70204	0.748363	-2.75855
8	-2.78252	0.991689	-0.67293
6	-2.83081	0.930493	0.682381
6	-4.13767	1.4721	1.21872
8	-1.908	0.500183	1.34539
6	-5.34191	0.698497	0.642041
6	-5.24923	-0.83741	0.727787
6	-6.51003	-1.46636	0.103601
6	-5.01215	-1.3188	2.16701
6	-4.23525	2.97921	0.918661
6	-6.4124	-2.97899	-0.12288
1	10.0219	0.861883	1.02621
1	9.78444	-0.5968	0.006194
1	9.47759	-0.68702	1.75349
1	-1.86965	1.81069	-2.94948
1	-2.55413	0.164679	-3.11348
1	-0.80108	0.422162	-3.28621
1	-4.9906	-2.41061	2.22314
1	-5.8082	-0.96452	2.83488
1	-4.05431	-0.96166	2.55624
1	-5.1559	3.38547	1.34882
1	-4.2481	3.15507	-0.1603
1	-3.38633	3.52528	1.3407
1	-7.29867	-3.35722	-0.64326
1	-5.53364	-3.22271	-0.73012
1	-6.32747	-3.52973	0.819164
1	-5.46021	0.98631	-0.40951
1	-6.24213	1.04551	1.16859
1	-6.70048	-0.97928	-0.86224
1	-7.38038	-1.23849	0.736557
1	-4.08389	1.32562	2.30077
1	-4.39419	-1.15298	0.115844
1	0.047924	-2.60425	-0.81956
1	2.55469	-2.27614	-0.41827
1	1.97342	2.57747	-0.27647
1	4.9545	-1.87597	0.005158
1	5.71658	1.09205	0.282637

Conformer 5: -1380.661154 hartrees

6	-1.70828	-1.20718	-1.044
6	-1.12066	-2.5505	-0.52456
6	0.316767	-2.63888	-0.33668
6	1.12181	-1.53681	-0.34376
6	0.599747	-0.19309	-0.55821
6	-0.846	0.041816	-0.78461
6	2.54702	-1.58687	-0.11521
6	3.3148	-0.46736	-0.11168
8	2.76371	0.782449	-0.32434
6	1.43357	0.870572	-0.52933
6	4.74316	-0.45144	0.10636
6	5.51166	0.651664	0.116227
6	6.96763	0.525948	0.352105
8	7.56038	1.74003	0.329591
8	7.56432	-0.51695	0.541498
6	8.98158	1.72773	0.551402
8	-1.3222	1.16282	-0.85634

8	-1.87601	-3.50042	-0.37132
6	-1.91367	-1.33244	-2.55985
8	-3.0199	-1.01853	-0.49685
6	-3.04852	-0.83734	0.847156
6	-4.45833	-0.65166	1.36568
8	-2.03566	-0.79038	1.51714
6	-4.77348	0.863851	1.43291
6	-4.66903	1.65075	0.111329
6	-4.92025	3.14662	0.387374
6	-5.61908	1.09458	-0.95946
6	-4.59634	-1.31114	2.74352
6	-4.59509	4.06869	-0.79281
1	9.29665	2.76957	0.502455
1	9.48449	1.13527	-0.21748
1	9.21301	1.29956	1.53026
1	-2.38161	-0.4191	-2.93419
1	-2.56102	-2.19071	-2.75188
1	-0.95718	-1.48515	-3.06813
1	-5.57012	1.68249	-1.88047
1	-6.65876	1.11124	-0.6076
1	-5.3616	0.065677	-1.22357
1	-5.60462	-1.15335	3.13773
1	-3.87404	-0.88278	3.44345
1	-4.41277	-2.38806	2.68718
1	-4.71353	5.12105	-0.51411
1	-3.56092	3.92231	-1.12436
1	-5.2485	3.88323	-1.65103
1	-4.09515	1.31017	2.17161
1	-5.79004	0.969011	1.83696
1	-4.30819	3.45048	1.24707
1	-5.9671	3.28843	0.693539
1	-5.138	-1.13152	0.656004
1	-3.64387	1.55535	-0.2667
1	0.7176	-3.6258	-0.13015
1	3.02866	-2.54208	0.062222
1	1.09511	1.88971	-0.66938
1	5.21619	-1.41532	0.274219
1	5.10413	1.64318	-0.04225

Conformer 6: -1380.660826 hartrees

6	1.72315	-1.62726	1.01081
6	1.24166	-2.62238	-0.08242
6	-0.16608	-2.619	-0.43605
6	-1.00892	-1.61352	-0.05807
6	-0.56392	-0.48662	0.75176
6	0.838572	-0.38436	1.22068
6	-2.3998	-1.55165	-0.44166
6	-3.2047	-0.52476	-0.06623
8	-2.72664	0.518295	0.70386
6	-1.42978	0.499758	1.0748
6	-4.60218	-0.40712	-0.41468
6	-5.40691	0.606813	-0.05153
6	-6.82459	0.603336	-0.47743
8	-7.46216	1.70327	-0.01965
8	-7.35838	-0.26105	-1.14638
6	-8.85109	1.79673	-0.38191
8	1.26131	0.588955	1.82173
8	2.04829	-3.41738	-0.54816
6	1.80693	-2.38374	2.34257
8	3.06546	-1.2171	0.716302
6	3.19106	-0.52486	-0.44267
6	4.6419	-0.30463	-0.81074
8	2.23111	-0.19526	-1.11171
6	4.86933	1.08758	-1.42272
6	4.94597	2.25717	-0.4196
6	3.68055	2.36787	0.452821
6	5.23967	3.55856	-1.18128
6	5.00999	-1.43674	-1.79549
6	3.69924	3.51074	1.47278
1	-9.20878	2.72677	0.058749
1	-9.40842	0.942826	0.012222
1	-8.96316	1.81602	-1.46917
1	2.19988	-1.71345	3.11027
1	2.47304	-3.23965	2.2149
1	0.818603	-2.74182	2.64538
1	5.42632	4.39555	-0.50293
1	4.39167	3.82928	-1.82237
1	6.1209	3.45106	-1.82334
1	6.08112	-1.40892	-2.01595
1	4.45828	-1.31408	-2.73249
1	4.75794	-2.4165	-1.37973
1	2.8286	3.4425	2.13228
1	4.59742	3.46547	2.10064
1	3.67856	4.49506	0.994432
1	4.06901	1.27588	-2.14954

1	5.80757	1.0558	-1.99064	1	-0.50645	-3.42507	-1.07773
1	3.53263	1.43619	1.00763	1	-2.82413	-2.34211	-1.05077
1	2.80314	2.46653	-0.19955	1	-1.14669	1.36374	1.6632
1	5.23904	-0.42073	0.100448	1	-5.01815	-1.20969	-1.01787
1	5.79589	2.06093	0.254076	1	-5.05606	1.43894	0.547458

Deflectin-1a (**3.10**)

Calculations of the ECD and UV spectra (B3LYP/6-31++G**) involved modeling the truncated (*R*) enantiomer of the natural product, shortening the exocyclic alkyl chain to a propyl group. This way there was still some conformational flexibility, but at a reduced computational cost. Since no other stereoisomers were possible, it should be noted that the (*S*)-enantiomer is assumed to have a spectrum that will be equal and opposite at all wavelengths.

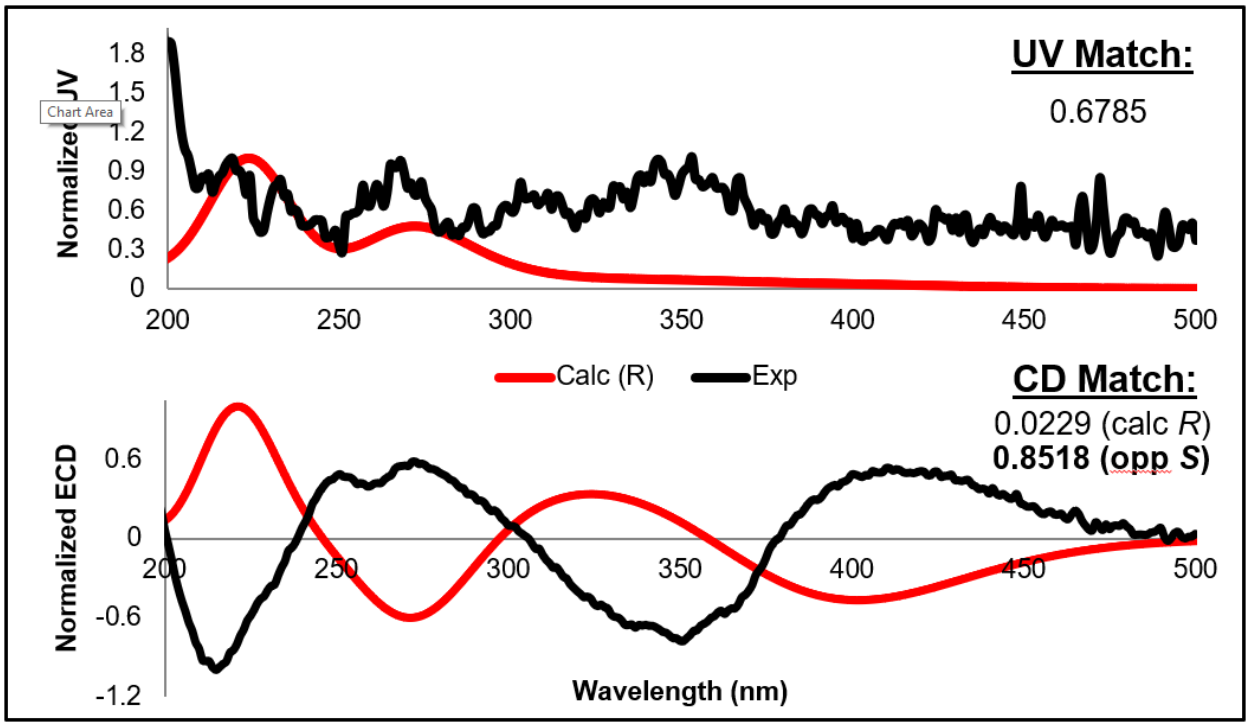


Figure 3.19. Comparison between experimental (black) and calculated (red) UV (top) and ECD (bottom) spectra. The calculated (*R*)-enantiomer is opposite of the experimental spectrum, with a large difference in fits (Δ=0.8289) suggesting a confident assignment. The calculated spectrum has been shifted -34 nm, and a band broadening of 0.3 eV has been applied.

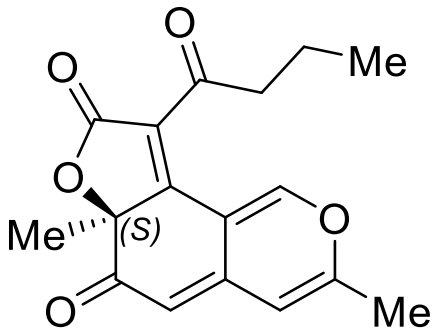


Figure 3.20. Assigned absolute configuration of truncated deflectin based on ECD analysis.

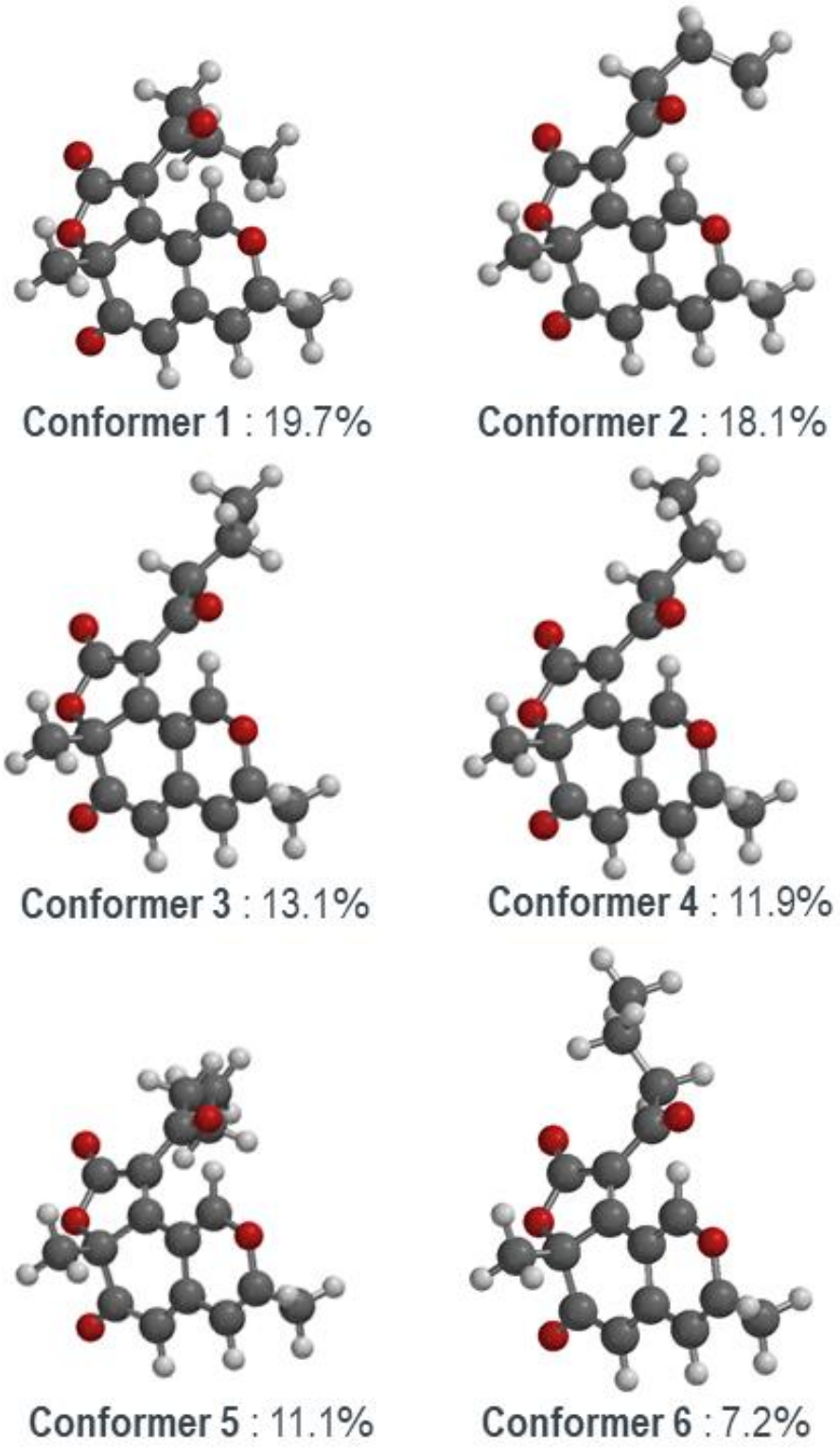


Figure 3.21. Six conformers of the (*R*)-enantiomer of truncated deflectin-1a that contribute >5% to the Boltzmann distribution. Note the percentage shown above based is on *in vacuo* electronic energies.

Coordinates and electronic energies for B3LYP/6-31G conformational minima contributing >5% to the *in vacuo* Boltzmann distribution.**

Conformer 1: -1033.613962 hartrees

6	1.12532	-0.04345	-0.39212
6	1.99275	0.732157	0.506105
6	3.30798	0.164545	0.702653
6	3.70807	-0.95125	0.058724
8	2.8718	-1.61446	-0.80807
6	1.61067	-1.16264	-0.98929
6	-0.24074	0.436799	-0.53748
6	-0.45225	1.9203	-0.3767
6	0.276203	2.47925	0.879938
6	1.57228	1.87365	1.13487
6	5.0355	-1.62273	0.154105
8	-0.18568	3.42614	1.49476
6	-1.45471	-0.16625	-0.63897
6	-2.49589	0.888061	-0.45779
8	-1.86713	2.0917	-0.28838
8	-3.69893	0.782672	-0.43922
6	-1.75682	-1.61382	-0.78063
8	-0.96298	-2.36216	-1.34426

Conformer 2: -1033.613885 hartrees

6	-1.12455	-0.26641	0.307283
6	-2.26771	0.221135	-0.47835
6	-3.34533	-0.7326	-0.62018
6	-3.31263	-1.94174	-0.02342
8	-2.23687	-2.33275	0.738887
6	-1.17513	-1.50542	0.861629
6	0.02309	0.624155	0.393856
6	-0.26797	2.1012	0.310862
6	-1.23803	2.44194	-0.8567
6	-3.0367	-2.1407	-0.16933
6	-3.06518	-1.968	1.36533
6	-1.88548	-2.64397	2.07088
6	0.047576	2.71521	-1.59817
1	4.91526	-2.65373	0.503621
1	5.69066	-1.08748	0.843001
1	5.51202	-1.66536	-0.83119
1	-1.96232	-2.53676	3.1572
1	-1.84539	-3.71359	1.8375
1	-0.92928	-2.20641	1.76156
1	-0.18359	3.77097	-1.44072
1	-0.46423	2.36183	-2.49698
1	1.12617	2.59644	-1.7238
1	-3.10279	-3.19773	-0.44239
1	-3.88139	-1.59557	-0.60229
1	-4.00926	-2.38602	1.73215
1	-3.09299	-0.90045	1.60707
1	4.00284	0.655146	1.37472
1	1.02743	-1.78942	-1.64834
1	2.20866	2.37403	1.8575
6	-2.28943	1.45998	-1.06108
6	-4.35993	-3.00144	-0.06963
8	-1.15626	3.51001	-1.44011
6	1.37198	0.449204	0.391607
6	1.99964	1.79257	0.218404
8	1.00322	2.72868	0.145072
8	3.16773	2.08818	0.142914
6	2.14281	-0.81834	0.425308
8	1.66299	-1.82714	0.93665
6	3.53106	-0.82161	-0.18453
6	4.14405	-2.21836	-0.30714
6	3.46052	-3.08584	-1.36975
6	-0.89721	2.64077	1.61047
1	-3.94811	-3.92427	-0.492

1	-5.20818	-2.68012	-0.67588
1	-4.71146	-3.23342	0.941464
1	3.94002	-4.06679	-1.44702
1	2.40785	-3.24503	-1.12034
1	3.51058	-2.61139	-2.35686
1	-1.0324	3.71945	1.5056
1	-0.22707	2.43922	2.4501
1	-1.86575	2.17166	1.79764

Conformer 3: -1033.613578 hartrees

6	-1.23669	0.367215	-0.25506
6	-2.39903	-0.12519	0.498848
6	-3.45192	0.849675	0.677926
6	-3.38244	2.08379	0.13827
8	-2.29045	2.48058	-0.59741
6	-1.2499	1.63178	-0.7509
6	-0.11296	-0.54982	-0.37714
6	-0.44715	-2.02001	-0.36658
6	-1.4339	-2.38875	0.778195
6	-2.45933	-1.38886	1.02254
6	-4.40172	3.16789	0.226283
8	-1.38534	-3.48506	1.31077
6	1.24071	-0.41422	-0.36279
6	1.82764	-1.78285	-0.25038
8	0.804028	-2.69148	-0.22508
8	2.98556	-2.11847	-0.18343
6	2.04843	0.831619	-0.33746
8	1.59408	1.88107	-0.78567

Conformer 4: -1033.613481 hartrees

6	-1.16263	-0.38101	0.220523
6	-2.40001	0.133099	-0.38523
6	-3.44651	-0.85331	-0.53626
6	-3.30464	-2.1192	-0.09307
8	-2.1423	-2.53751	0.51102
6	-1.10736	-1.67595	0.626863
1	-4.14471	-0.15612	0.43636
1	-3.49183	-0.31005	-1.15492
1	-4.08734	-2.71781	0.665863
1	-5.20723	-2.10655	-0.54866
1	-4.21514	-0.46655	-1.20995
1	-0.36813	-1.93743	1.43516
1	-3.11123	1.75434	-1.70571
6	3.43612	0.767733	0.264145
6	4.12596	2.12836	0.346454
6	5.53042	2.02448	0.946302
6	-1.08342	-2.47786	-1.69396
1	-3.96898	4.05911	0.693254
1	-5.26285	2.84194	0.811555
1	-4.73918	3.45517	-0.77537
1	6.01292	3.00558	0.997751
1	5.49985	1.6137	1.96193
1	6.17009	1.36781	0.346005
1	-1.25056	-3.55592	-1.64194
1	-0.40215	-2.25596	-2.5194
1	-2.03663	-1.97206	-1.86332
1	4.01855	0.049544	-0.32738
1	3.36471	0.290609	1.25149
1	3.50852	2.81195	0.940619
1	4.17124	2.56697	-0.65634
1	-4.33325	0.579987	1.24865
1	-0.42639	2.06877	-1.29639
1	-3.29392	-1.69019	1.64721
6	-0.04957	0.55392	0.301997
6	-0.41246	2.01128	0.432348
6	-1.51029	2.43529	-0.58532
6	-2.53386	1.42861	-0.80765
6	-4.30737	-3.21985	-0.16391
8	-1.53384	3.56825	-1.03667
6	1.29868	0.455493	0.146455
6	1.84474	1.84502	0.081729
8	0.80578	2.72429	0.224366

8	2.98285	2.21887	-0.06994
6	2.11868	-0.76834	-0.04675
8	1.72595	-1.85224	0.377492
6	3.42359	-0.63768	-0.80651
6	4.20384	-1.94904	-0.92925
6	4.84226	-2.39582	0.390592
6	0.92807	2.35529	1.84364
1	3.90327	-4.06598	-0.72992
1	5.22487	-2.87623	-0.64395
1	4.54526	-3.58327	0.841596
1	5.41245	-3.32099	0.258882
1	5.52894	-1.63143	0.772205
1	4.07651	-2.57643	1.14925
1	1.12083	3.42941	1.88756
1	0.16683	2.09296	2.58284
1	-1.8505	1.81319	2.06358
1	4.01663	0.149602	-0.32628
1	3.17871	-0.22008	-1.79379
1	4.98355	-1.81297	-1.68758
1	3.53375	-2.73376	-1.29737
1	4.38175	-0.5671	-1.00381
1	0.22974	-2.12859	1.06416
1	3.42927	1.75158	-1.32867

Conformer 5: -1033.613417 hartrees

6	-1.28767	0.218168	-0.29924
6	-2.21665	-0.5007	0.585253
6	-3.45445	0.199085	0.84786
6	-3.74166	1.38401	0.271162
8	-2.8556	1.99399	-0.58535
6	-1.65796	1.41507	-0.82441
6	0.010014	-0.40554	-0.5129
6	0.047735	-1.9102	-0.42909
6	-0.70447	-2.44705	0.822674
6	-1.91543	-1.71289	1.14633
6	-4.98261	2.19305	0.436761
8	-0.33731	-3.47014	1.37615
6	1.28483	0.053474	-0.64059
6	2.19597	-1.1267	-0.54633
8	1.43292	-2.25264	-0.40061
8	3.40297	-1.17042	-0.57187
6	1.75903	1.45745	-0.75213
8	1.01665	2.33155	-1.19275
6	3.16063	1.78466	-0.28827
6	3.34581	1.5836	1.23018
6	4.76986	1.9244	1.67516
6	-0.57977	-2.57506	-1.67001
1	-4.7392	3.18535	0.831304
1	-5.6774	1.69967	1.11806
1	-5.47437	2.33802	-0.53116
1	4.88825	1.7911	2.75506
1	5.49623	1.2759	1.17413
1	5.02313	2.96293	1.4345
1	-0.4718	-3.6575	-1.57239
1	-0.05535	-2.23779	-2.56775
1	-1.63953	-2.32352	-1.75203
1	3.35205	2.82513	-0.56546
1	3.86638	1.13206	-0.81272
1	3.12583	0.542541	1.48743
1	2.62291	2.2093	1.76876
1	-4.18405	-0.24573	1.51496
1	-1.02206	2.00861	-1.46489
1	-2.58722	-2.17362	1.86324

Conformer 6: -1033.613007 hartrees

6	-1.33004	-0.33625	0.191808
6	-2.46505	0.40933	-0.37277
6	-3.67628	-0.36659	-0.52076
6	-3.75735	-1.64969	-0.11357
8	-2.67727	-2.28908	0.448706
6	-1.50175	-1.63261	0.561122
6	-0.06485	0.378031	0.271432
6	-0.15474	1.8741	0.43505
6	-1.17503	2.515	-0.54881
6	-2.36977	1.71762	-0.7651
6	-4.94413	-2.5486	-0.1861
8	-0.99925	3.64314	-0.97797
6	1.24305	0.040367	0.096574
6	2.0288	1.30911	0.044704

8	1.16892	2.35948	0.216565
8	3.21456	1.47291	-0.11853
6	1.83025	-1.30289	-0.13255
8	1.2409	-2.31551	0.238784
6	3.18478	-1.41488	-0.79951
6	4.30321	-1.49201	0.263832
6	5.68656	-1.61045	-0.37837
6	-0.57534	2.27283	1.86396
1	-4.71207	-3.43639	-0.78408
1	-5.79421	-2.0308	-0.63283
1	-5.22243	-2.89313	0.815677
1	6.46885	-1.68168	0.384072
1	5.75537	-2.49987	-1.01506
1	5.90155	-0.73386	-0.99846
1	-0.5682	3.36281	1.93305
1	0.138091	1.85931	2.58141
1	-1.57714	1.90262	2.09278
1	3.37509	-0.5662	-1.46007
1	3.1723	-2.33988	-1.38505
1	4.10728	-2.35242	0.914763
1	4.26382	-0.59047	0.882688
1	-4.55345	0.097571	-0.95702
1	-0.7104	-2.24947	0.96088
1	-3.20125	2.20988	-1.25887

3.8 References

(1) Sun, A. W.; Lackner, S.; Stoltz, B. M., Modularity: Adding New Dimensions to Total Synthesis. *Trends Chem.* **2019**, *1*, 630-643.

(2) Stark, L. M.; Pekari, K.; Sorensen, E. J., A nucleophile-catalyzed cycloisomerization permits a concise synthesis of (+)-harziphilone. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12064-12066.

(3) Chong, R.; Gray, R. W.; King, R. R.; Whalley, W. B., The synthesis of (±) mitorubrin. *J. Chem. Soc. D* **1970**, *2*, 101a-101a.

(4) Marsini, M. A.; Gowin, K. M.; Pettus, T. R. R., Total Synthesis of (±)-Mitorubrinic Acid. *Org. Lett.* **2006**, *8*, 3481-3483.

(5) Zhu, J.; Germain, A. R.; Porco Jr., J. A., Synthesis of Azaphilones and Related Molecules by Employing Cycloisomerization of o-Alkynylbenzaldehydes. *Angew. Chem. Int. Ed.* **2004**, *43*, 1239-1243.

(6) Makrerougras, M.; Coffinier, R.; Oger, S.; Chevalier, A.; Sabot, C.; Franck, X., Total Synthesis and Structural Revision of Chaetoviridins A. *Org. Lett.* **2017**, *19*, 4146-4149.

(7) Kang, H.; Torruellas, C.; Liu, J.; Kozlowski, M. C., Total Synthesis of Chaetoglobin A via Catalytic, Atroposelective Oxidative Phenol Coupling. *Org. Lett.* **2018**, *20*, 5554-5558.

(8) Zhu, J.; Grigoriadis, N. P.; Lee, J. P.; Porco, J. A., Synthesis of the Azaphilones Using Copper-Mediated Enantioselective Oxidative Dearomatization. *J. Am. Chem. Soc.* **2005**, *127*, 9342-9343.

(9) Zhu, J.; Porco, J. A., Asymmetric Syntheses of (−)-Mitorubrin and Related Azaphilone Natural Products. *Org. Lett.* **2006**, *8*, 5169-5171.

(10) Germain, A. R.; Bruggemeyer, D. M.; Zhu, J.; Genet, C.; O'Brien, P.; Porco, J. A., Synthesis of the azaphilones (+)-sclerotiorin and (+)-8-O-methylsclerotiorinamine utilizing (+)-sparteine surrogates in copper-mediated oxidative dearomatization. *J. Org. Chem.* **2011**, *76*, 2577-2584.

(11) Peixoto, P. A.; Boulangé, A.; Ball, M.; Naudin, B.; Alle, T.; Cosette, P.; Karuso, P.; Franck, X., Design and Synthesis of Epicocconone Analogues with Improved Fluorescence Properties. *J. Am. Chem. Soc.* **2014**, *136*, 15248-15256.

(12) Davison, J.; al Fahad, A.; Cai, M.; Song, Z.; Yehia, S. Y.; Lazarus, C. M.; Bailey, A. M.; Simpson, T. J.; Cox, R. J., Genetic, molecular, and biochemical basis of fungal tropolone biosynthesis. *Proc. Natl. Acad. Sci.* **2012**, *109*, 7642-7647.

(13) Thines, E.; Anke, H.; Sterner, O., Trichoflectin, a Bioactive Azaphilone from the Ascomycete Trichopezizella nidulus. *J. Nat. Prod.* **1998**, *61*, 306-308.

(14) Steyn, P. S.; Vleggaar, R., The structure of dihydrodeoxy-8-epi-austdiol and the absolute configuration of the azaphilones. *J. Chem. Soc., Perkin Trans. 1* **1976**, *2,* 204-206.

(15) Anke, H.; Kemmer, T.; Höfle, G., Deflectins, new antimicrobial azaphilones from Aspergillus deflectus. *J. Antiobiot.* **1981**, *34*, 923-928.

(16) Nukina, M.; Marumo, S., Lunatoic acid A and B, aversion factor and its related metabolite of cochliobolus lunata. *Tetrahedron Lett.* **1977**, *18*, 2603-2606.

(17) Marumo, S.; Nukina, M.; Kondo, S.; Tomiyama, K., Lunatoic Acid A, a Morphogenic Substance Inducing Chlamydospore-like Cells in Some Fungi. *Agricultural and Biological Chemistry* **1982**, *46* (9), 2399-2401.

(18) Kawanami, Y.; Ito, Y.; Kitagawa, T.; Taniguchi, Y.; Katsuki, T.; Yamaguchi, M., Asymmetric alkylation of carboxyamides by using trans-2,5-disubstituted pyrrolidines as chiral auxiliaries. *Tetrahedron Lett.* **1984**, *25*, 857-860.
(19) Myers, A. G.; Yang, B. H.; Chen, H.; McKinstry, L.; Kopecky, D. J.; Gleason, J. L., Pseudoephedrine as a Practical Chiral Auxiliary for the Synthesis of Highly Enantiomerically Enriched Carboxylic Acids, Alcohols, Aldehydes, and Ketones. *J. Am. Chem. Soc.* **1997**, *119*, 6496-6511.
(20) Blackwell, H. E.; O'Leary, D. J.; Chatterjee, A. K.; Washenfelder, R. A.; Bussmann, D. A.; Grubbs, R. H., New Approaches to Olefin Cross-Metathesis. *J. Am. Chem. Soc.* **2000**, *122*, 58-71
(21) Frigerio, M.; Santagostino, M.; Sputore, S., A User-Friendly Entry to 2-Iodoxybenzoic Acid (IBX). *J. Org. Chem.* **1999** *64*, 4537-4538.
(22) Baker Dockrey, S. A.; Lukowski, A. L.; Becker, M. R.; Narayan, A. R. H., Biocatalytic site- and enantioselective oxidative dearomatization of phenols. *Nat. Chem.* **2018**, *10*, 119-125.
(23) MichaelChong, J.; Shen, L., Preparation of Chloromethyl Methyl Ether Revisited. *Synth. Commun.* **1998**, *28*, 2801-2806.
(24) Evans, D. A.; Fandrick, K. R.; Song, H.-J., Enantioselective Friedel−Crafts Alkylations of α,β-Unsaturated 2-Acyl Imidazoles Catalyzed by Bis(oxazolinyl)pyridine−Scandium(III) Triflate Complexes. *J. Am. Chem. Soc.* **2005**, *127*, 8942-8943.
(25) Fuse, S.; Yoshida, H.; Oosumi, K.; Takahashi, T., Rapid and Structurally Diverse Synthesis of Multi-Substituted β-Keto Amide Derivatives Based on a Dioxinone Scaffold. *Eur. J. Org. Chem.* **2014**, *2014*, 4854-4860.
(26) Dolomanov, O. V.; Bourhis, L. J.; Gildea, R. J.; Howard, J. A. K.; Puschmann, H., OLEX2: a complete structure solution, refinement and analysis program. *J. Appl. Crystallogr.* **2009**, *42*, 339-341.
(27) Sheldrick, G. M., SHELXT - integrated space-group and crystal-structure determination. *Acta Crystallogr. A* **2015**, *71*, 3-8.
(28) Sheldrick, G. M., Crystal structure refinement with SHELXL. *Acta Crystallogr. C* **2015**, *71*, 3-8.
(29) Sherer, E. C.; Lee, C. H.; Shpungin, J.; Cuff, J. F.; Da, C.; Ball, R.; Bach, R.; Crespo, A.; Gong, X.; Welch, C. J., Systematic approach to conformational sampling for assigning absolute configuration using vibrational circular dichroism. *J. Med. Chem.* **2014**, *57*, 477-94.
(30) Liu, Z.; Shultz, C. S.; Sherwood, C. A.; Krska, S.; Dormer, P. G.; Desmond, R.; Lee, C.; Sherer, E. C.; Shpungin, J.; Cuff, J.; Xu, F., Highly enantioselective synthesis of anti aryl β-hydroxy α-amino esters via DKR transfer hydrogenation. *Tetrahedron Lett.* **2011**, *52*, 1685-1688.
(31) Joyce, L. A.; Nawrat, C. C.; Sherer, E. C.; Biba, M.; Brunskill, A.; Martin, G. E.; Cohen, R. D.; Davies, I. W., Beyond optical rotation: what's left is not always right in total synthesis. *Chem. Sci.* **2018**, *9*, 415-424.
(32) Petersson, G. A.; Al-Laham, M. A., A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms. *J. Chem. Phys.* **1991**, *94*, 6081-6090.

(33) Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J., A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements. *J. Chem. Phys.* **1988**, *89*, 2193-2218.
(34) Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L., 6-31G* basis set for atoms K through Zn. *J. Chem. Phys.* **1998**, *109*, 1223-1229.
(35) Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A., 6-31G* basis set for third-row atoms. *J. Comput. Chem.* **2001**, *22*, 976-984.
(36) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A., Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654-3665.
(37) Hehre, W. J.; Ditchfield, R.; Pople, J. A., Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257-2261.
(38) Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
(39) Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785-789.
(40) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H., Results obtained with the correlation energy density functionals of becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **1989**, *157*, 200-206.
(41) M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox.
(42) Yanai, T.; Tew, D. P.; Handy, N. C., A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51-57.
(43) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378-6396.
(44) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R., Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F. *J. Comput. Chem.* **1983**, *4*, 294-301.

(45) Bauernschmitt, R.; Ahlrichs, R., Treatment of electronic excitations within the adiabatic approximation of time dependent density functional theory. *Chem. Phys. Lett.* **1996**, *256*, 454-464.
(46) Bruhn, T.; Schaumlöffel, A.; Hemberger, Y.; Bringmann, G., SpecDis: Quantifying the Comparison of Calculated and Experimental Electronic Circular Dichroism Spectra. *Chirality* **2013**, *25*, 243-249.
(47) Mazzeo, G.; Santoro, E.; Andolfi, A.; Cimmino, A.; Troselj, P.; Petrovic, A. G.; Superchi, S.; Evidente, A.; Berova, N., Absolute Configurations of Fungal and Plant Metabolites by Chiroptical Methods. ORD, ECD, and VCD Studies on Phyllostin, Scytolide, and Oxysporone. *J. Nat. Prod.* **2013**, *76*, 588-599.

Chapter 4: One-pot synthesis of tricyclic azaphilones using a two enzyme sequence

Pyser, J. B.; Wang, Y.; and Narayan, A. R. H. *Manuscript in preparation.*

Summary

Polyketide synthases and their related tailoring enzymes can construct and decorate natural product scaffolds with ease and efficiency. Acyltransferases (ATs) are a common enzyme domain from pathways containing PKSs that can exist as part of the PKS or function independently post-PKS, such as in the azaphilone natural product class. Despite their abundance and critical role in biosynthesis, there are a limited number of reports describing the use of ATs in biocatalysis and natural product total synthesis. To streamline the synthesis of tricyclic azaphilone natural products and their analogs, we describe the use of the AT MrPigD in a one-pot enzymatic sequence to provide access to these scaffolds from resorcinol substrates. Combining this enzyme with the FAD-dependent monooxygenase (FDMO) AzaH and a pantetheine-based acyl group donor allowed for the conversion of several orcinaldehyde substrates directly to the corresponding tricyclic azaphilone in a single reaction vessel, eliminating the need to isolate the dearomatized intermediate. This approach was used to provide preparative-scale quantities of three azaphilone tricycles, including the natural product rubropunctatin (**4.10**). NMR characterization of these compounds indicated that the acylated

intermediate can undergo a subsequent cyclization reaction with selectivity that is complementary to those reported previously, providing access to linear tricyclic scaffolds with unprecedented selectivity.

4.1 Introduction

Biosynthetic pathways consist of enzymatic machinery responsible for constructing complex natural products with valuable biological activity.[1-5] Metabolites from these pathways can exhibit impressive structural diversity in addition to interesting pharmacological properties, ranging from enormous marine toxins,[6] large macrocyclic antibiotics,[7] and functionalized bicyclic protein inhibitors.[4] The synthetic chemistry community now recognizes the operational efficiency of enzymes like the ones responsible for these biosyntheses; the field as a whole continues to embrace the use of enzymes in total synthesis.[8-18] With recent advances in biocatalysis and molecular biology lowering the barrier for their use by making it easier and less expensive to obtain them,[10-13, 15, 19-24] there is incredible potential to further exploit catalytic proteins like those from natural product biosynthetic pathways for synthetic endeavors.

Azaphilones are one class of natural products biosynthesized by fungi through polyketide synthase (PKS) pathways.[25-27] This large natural product class exhibits impressive structural diversity, made possible by the seamless cooperation of the PKS

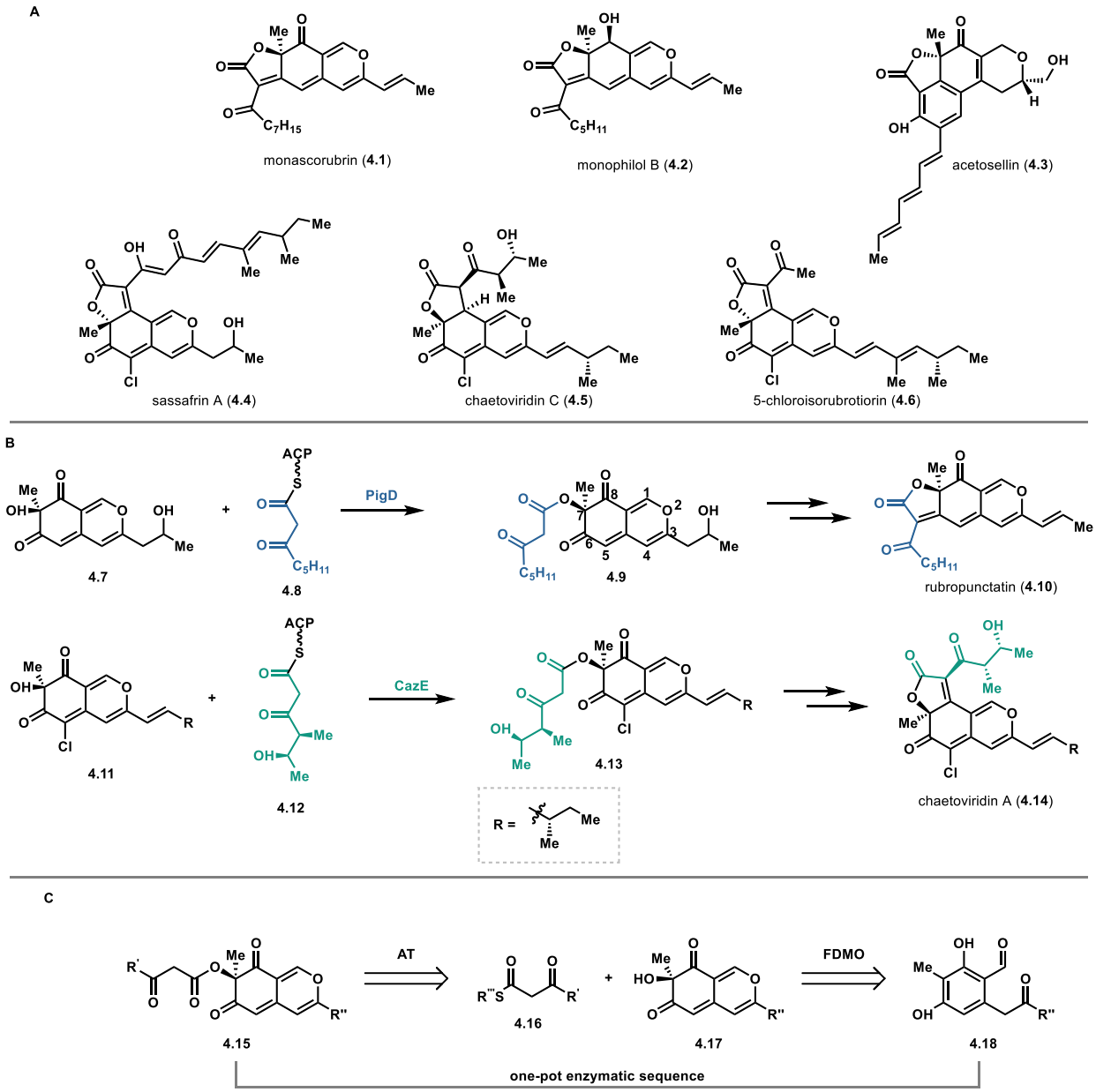


Figure 4.1. A) Azaphilone natural products containing additional ring systems. B) Native reactions of ATs PigD and CazE. C) Biocatalytic retrosynthesis of acylated azaphilones.

architecture and downstream tailoring enzymes.[27] In particular, many azaphilones with potential medicinal properties have been identified that bear additional rings such as butenolides or lactones appended to the basic pyranoquinone core (Figure 4.1A).[27-31] Typically, these ring systems are constructed following introduction of the acyl group by acyltransferases (ATs) operating in conjunction with other substrates and domains from

the pathway.[25, 32] Such as in the biosynthesis of azaphilone natural products rubropunctatin (**4.10**) and chaetoviridin A (**4.14**), the ATs MrPigD[25] (PigD) and CazE[32] catalyze the transfer of an ACP-bound fatty acyl chain to the C7 hydroxyl on the azaphilone bicycle (Figure 4.1B). Following acylation, the linear intermediate is reported to undergo a spontaneous Knoevenagel condensation, with the newly introduced acyl chain forming a C–C bond with either the C6 or C8 carbonyl group carbon, resulting in a linear or angular tricycle, respectively.[25, 32-35] The pharmacology of these highly functionalized azaphilones remains poorly understood,[29, 30, 36, 37] motivating the development of better synthetic techniques to diversify these scaffolds for examination in biological assays and structure activity relationship (SAR) studies.

ATs are commonly found as part of biosynthetic pathways and play a vital role in adding complexity to natural products cores constructed by PKSs.[38-40] Despite being essential in the biosynthesis of many metabolites,[41] relatively few recent reports describe the use of ATs for *in vitro* biocatalysis. Schmidt and coworkers engineered a hetero-trimeric bacterial AT construct, called *PpATaseCH,* as a biocatalyst to efficiently affect two types of acylation reactions on several resorcinol substrates.[42] Furthermore, this enzyme accepts a variety of acyl donor substrates to affect the transformation, demonstrating impressive substrate promiscuity and eliminating the need to rely on a CoA-activated fatty acid substrates to load the AT.[42] Xie and coworkers also reported the use of LovD, an AT from the lovastatin biosynthetic pathway, as a biocatalyst.[39] This enzyme is capable of generating a diverse array of lovastatin analogs *in vitro* by regioselective acylation of the C8 hydroxyl group on monacolin J and 6-hydroxyl-6-desme-thylmonacolin J.[39] Wild type LovD exhibits impressive promiscuity towards the

acyl-donor reagent, accepting a total of 22 different thioester substrates and thus removing the need for an ACP for productive catalysis.[39]

Previously, our group[35] and others[33, 34] have reported the total synthesis of tricyclic azaphilones natural products that possess a third ring most likely installed from the action of ATs. However, these syntheses require forcing reaction conditions and isolation of reactive dearomatized intermediates, in addition to producing mixtures of linear (such as **4.1**) and angular (see **4.4**) tricyclic products depending on the particular azaphilone substrate used.[33-35] For example, in their synthesis of analogs of the linear tricyclic azaphilone natural product epicocconone, Franck and coworkers report some degree of regioselectivity in favor of linear tricycle production by leveraging steric effects from bulky acyl groups.[34] Despite this important contribution to the field in synthesizing the first synthesis of a linear tricyclic azaphilone, the drawback to this methodology is that it is limited to specific bulky acyl groups to achieve the linear selectivity. It is theorized that the increased steric bulk afforded by these larger substituents is responsible for biasing the condensation regioselectivity towards the linear scaffold.[34] By contrast, Nature is able to achieve the condensation regioselectivity affording the linear tricyclic scaffold through manipulations of the azaphilone core π-system.[25] Such is the case for the natural product rubropunctatin (**4.10**). Following acylation by PigD, the substrate is reported to undergo enzymatic reduction, truncating the conjugated π system of the bicyclic intermediate. This reportedly changes the activity of the two ketones such that the C6 carbonyl is more now electrophilic than the C8 carbonyl, allowing for cyclization in the linear direction. The initial π-system is then restored by a downstream oxidoreductase.[25]

We envisioned an improved synthesis could proceed through a one-pot, two-enzyme sequence to afford linear tricyclic azaphilones. Given that enzymes can operate synergistically in Nature,[5] we sought to identify an AT to provide efficient access to acylated azaphilones in a biomimetic approach (Figure 4.1C). To access analogs of acylated azaphilone bicycles (**4.15**), we investigated the AT PigD from a *Monascus ruber* biosynthetic pathway[25] for its ability to acylate azaphilone intermediates (i.e. **4.17**). Inspired by the efficiency of biocatalysts like LovD operating without the putative ACP,[39] we sought to substitute the canonical *Monascus* ACP from this pathway with a synthetic thioester **4.16** as shown previously,[25] eliminating the need for a protein-bound acyl substrate in our system. To further streamline the synthesis of azaphilone tricycles, we proposed constructing the dearomatized acyl-acceptor substrates (**4.17**) *in situ* through the dearomatization of orcinaldehyde starting materials (such as **4.18**) by AzaH[35, 43] (Figure 4.1C).

4.2 Selecting and Optimizing Acyltransferase PigD

Studies of azaphilone biosynthesis have revealed many pathways containing ATs performing acylations on dearomatized compounds.[1] After a careful search of relevant literature, we identified PigD[25] and CazE[32] as potential enzymes for investigation. Performing similar roles in the production of tricyclic azaphilones, these enzymes were chosen due to their reported promiscuity (catalyzing the transfer of several different acyl chains), the established precedent for their production in *E. coli*, and their ability to perform their reported functions *in vitro*.[25, 32] Chen and coworkers also report that PigD recognizes a pantetheine-based thioester (**4.20**, Figure 4.3A) as the acyl donor in place of the native ACP, further reinforcing our interest in this enzyme.[25] In addition, to provide the best opportunity of identifying an AT with suitable activity and substrate promiscuity, a sequence similarity network (SSN)[44-46] with an alignment score of 90 was constructed using the *pigD* gene as the input sequence (Figure 4.2). From this SSN, 12 putative AT sequences with varying degrees of relatedness to PigD were identified and ten were selected for investigation. These particular sequences were chosen through a subsequent genome neighborhood analysis

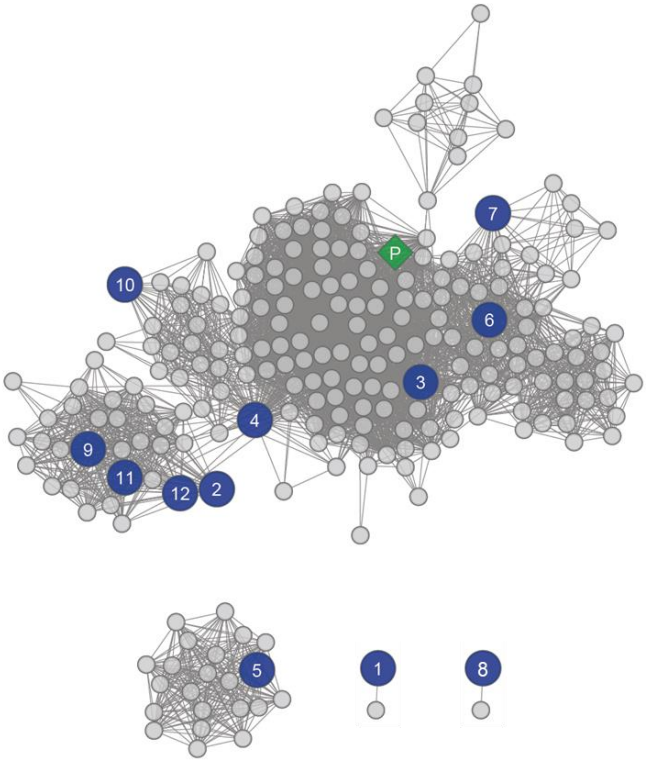


Figure 4.2. SSN of transferase proteins related to PigD created using web tools originating from the Enzyme Function Initiative**.** The green diamond represents the *pigD* sequence and the blue circles represent *tran1-12* sequences. This SSN was generated with an alignment score of 90.

indicating that, similarly to PigD and CazE, they also neighbor an FDMO in their respective pathways,.[44, 45]

To test the hypothesis that an AT could be used to functionalize dearomatized intermediates *in vitro*, the synthetic genes for *pigD*, *cazE*, and the ten uncharacterized ATs were transformed into *E. coli* BL21(DE3) cells and the proteins were expressed following reported conditions.[25, 32] Enantioenriched azaphilone **4.19** was prepared as reported previously[43] and added to clarified cell lysates containing the ATs (see Experimental) under the conditions reported in Figure 4.3A. No product was detected in the samples containing CazE and the ten uncharacterized ATs, most likely due to low levels of expression (Figure 4.6). The reaction containing PigD lysate however adopted a distinct red color following incubation for 30 min at 30 °C. Analysis of the reaction by LC-MS revealed a mass corresponding to successful acylation. Analysis of this reaction by UPLC with monitoring at 450 nm also suggested formation of a tricyclic azaphilone: a

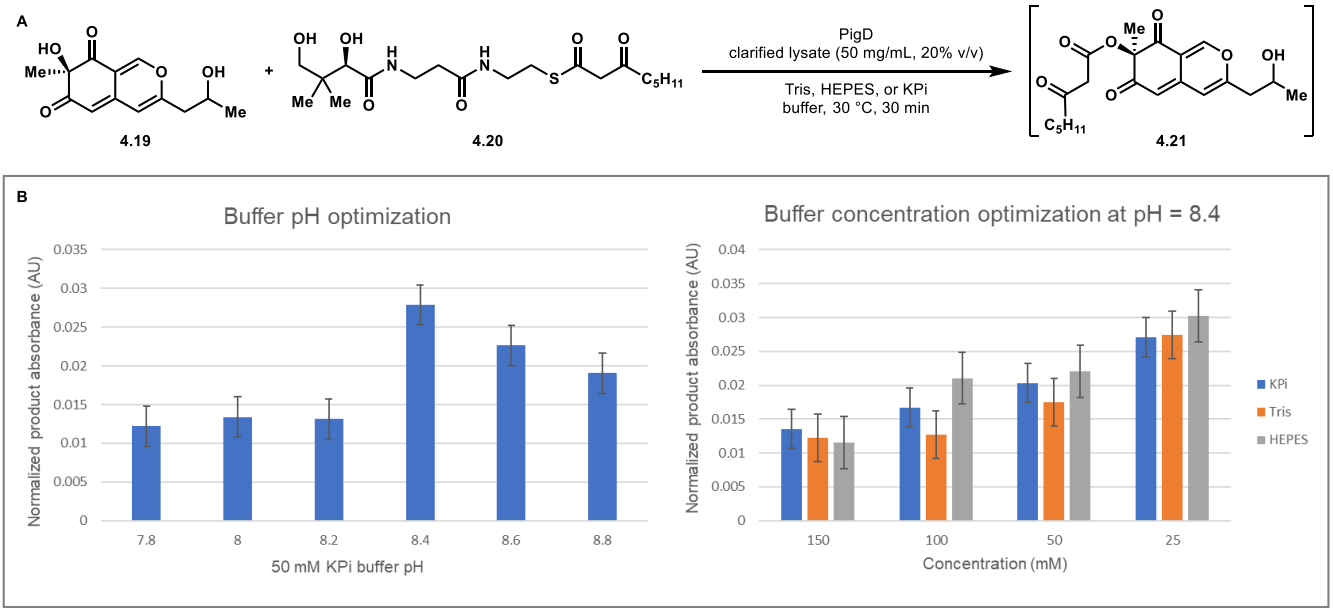


Figure 4.3. A) Reaction scheme of PigD clarified lysate with native azaphilone bicycle. B) Buffer pH and concentration optimization data. Product was detected at 450 nm and normalized to the pentamethyl benzene internal standard (1 mM final concentration) at 270 nm. All reactions were performed in duplicate.

UV signature matching that reported by Chen and coworkers in PigD *in vitro* reactions was detected (see Experimental). Following this initial success, a brief optimization of reaction conditions was performed (Figure 4.3B). These experiments indicate that PigD is tolerant of varied conditions, but optimal production of **4.21** appears to occur in 25 mM Tris, KPi, or HEPES buffer pH 8.4.

Upon reaction quenching, we observed significant red coloration of the insoluble material following centrifugation, suggesting the product molecule may have precipitated alongside the protein through adherence to the biological material or through a covalent attachment to the protein itself. As these tricyclic azaphilones are known to readily condense onto amino acids through formation of a Schiff base,[25, 30, 34, 47] we suspected any free amines from remaining endogenous *E. coli* proteins in the cell lysate could be contributing to precipitation of **4.21**. Therefore, purified PigD was obtained using a protocol modified from a previous report[25] (see Experimental) and used in lieu of clarified cell lysate for the remainder of this study.

4.3 Substrate Screen and Preparative-Scale Reactions with PigD

With optimized conditions and purified PigD in hand, we then investigated the ability of this enzyme to functionalize dearomatized intermediates formed *in situ* using AzaH (Figure 4.4A). First, a focused panel of orcinaldehyde substrates examined previously was dearomatized using AzaH under standard conditions.[35, 43] After confirming successful dearomatization, the pH of the reactions was adjusted from 8 to 8.4, 1 equivalent of thioester **4.20** was added, and 20 μM PigD was added to the mixture. Gratifyingly, a significant color change was observed with the five substrates depicted in

Figure 4.4B under these conditions. UPLC-DAD and LC-MS analysis of these reactions confirmed the presence of both the acylated and tricyclic material, whereas the AzaH only negative controls did not contain masses corresponding to the desired tricyclic products (see Experimental). At 450 nm, we observed a peak in the UPLC trace from four of the five reactions with a UV signature matching that observed for the product following acylation of **4.19** (see Experimental). This peak was not present in the AzaH only controls, supporting the hypothesis that PigD is indeed responsible for mediating this transformation (see Experimental). Additionally, both the $[M+H]^+$ and $[M+H+H_2O]^+$ product masses can be seen for reactions containing PigD compared to spectra collected for the AzaH only samples (see Experimental).

Despite an obvious color change (C, Figure 4.4B), the UV signature of the expected peak from this sample did not resemble that of **4.25A**. Instead, this spectrum

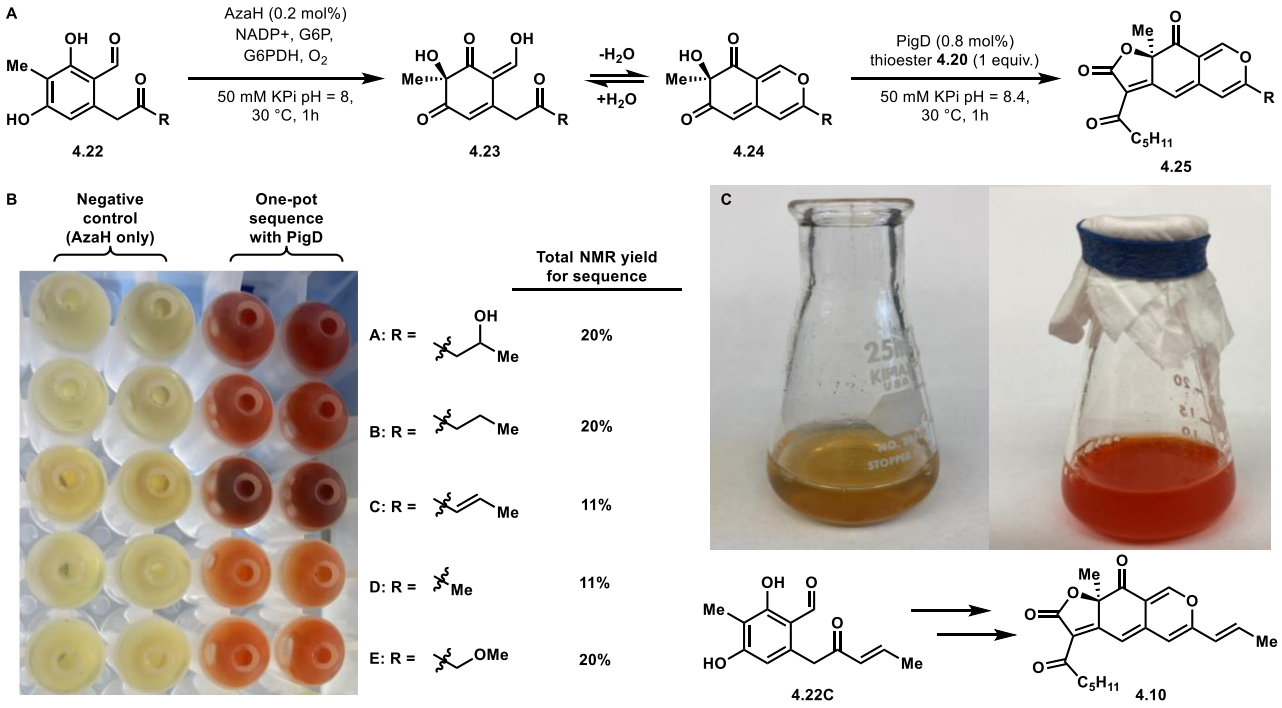


Figure 4.4. A) Reaction scheme of the one-pot transformation of orcinaldehydes by AzaH and PigD. B) Substrate screen of PigD in a well plate. C) Preparative scale one-pot synthesis of rubropunctatin (**4.10**).

matches those publishIed for the natural products rubropunctatin and monascorubrin (see Experimental).[25] This result was unexpected, as it indicated that the acylated intermediate had undergone a spontaneous Knoevenagel condensation onto the C6 carbonyl of the azaphilone core to furnish a linear tricycle. This cyclization selectivty conflicts with that observed from our previous experiments with this particular substrate in our total synthesis of the enantiomers of the angular tricyclic azaphilone trichoflectin (**3.1**, Chapter 3),[35] and it contradicts the report describing how the biosynthetic pathway of rubropunctatin (**4.10**) and monascorubrin (**4.1**) achieves this selectivity.[25]

To confirm that enzymatic acylation led to formation of the linear tricycle to produce the natural product rubropunctatin (**4.10**), a prepartive-scale reaction was performed to provide material for NMR studies (Figure 4.4C). 2 mg of the enone orcinaldehyde (**4.22C**, Figure 4.4C) was first dearomatized with AzaH, followed by a pH adjustment and addition of **4.20** and PigD to the reaction mixture. After one hour, the reaction was diluted with saturated Na_2SO_4. The crude material was then extracted from the aqueous phase by liquid-liquid extraction with 9:1 ethyl acetate:*n*-butanol before it was concentrated and subjected to separation conditions by preperative HPLC (see Experimental). The ^{1}H and ^{13}C NMR spectra from the purified material, isolated as an amorphous orange solid, matches with those reported for rubropunctatin and monascorubrin, indicating the succesful total synthesis of **4.10**.[48]

After confirming the condensation regioselectivity of this transformation, we investigated whether this observation held true for other substrates. To access material for additional NMR studies, we performed a preparative scale reaction with the reported native substrate of PigD (**4.22A**, Figure 4.4B) and an orcinaldehyde containing a propyl

ketone (**4.22B**, Figure 4.4B), under the same conditions. Following separation and purification, comparison of the resulting ^{1}H spectra to that of **4.10** shows an excellent correlation, particularly those of the signals afforded by the vinylogous protons appended to C1, C4, and C5 on the azaphilone pyranoquinone core (see Experimental). This strongly suports the selective formation of linear tricycles under these conditions and suggests that the preferance for the C6 carbonyl during the condensation reaction is not isolated to a single substrate, as the characteristic vinylogous proton signals are signifiacntly shifted in NMR spectra published for similar angular tricycles.[33-35]

The mechanism for this observed cyclization selectivty is currently not understood; however we have developed two working hypotheses. The first being that the preferance for condensation onto the C6 carbonyl is controlled directly by PigD through an unknown substrate-enzyme interaction. Alternativley, we suspect that the true substrate for PigD is not the “closed” pyranoquinone (**4.24**, Figure 4.4A) formed from a dehydration event following dearomatization of the aromatic substrate by an FDMO.[1, 25, 43] Instead, we speculate that PigD perfers the “open” dienol form of the dearomatized intermediate (**4.23**, Figure 4.5) as its substrate. This theory is supported by the observation that the dienol is the direct product of FDMOs like AzaH and that the dearomatized species are persistent in solution.[43] We have also observed only trace product formation when PigD is used to acylate the isolated bicycle directly as opposed to when this dearomatized material is generated *in*

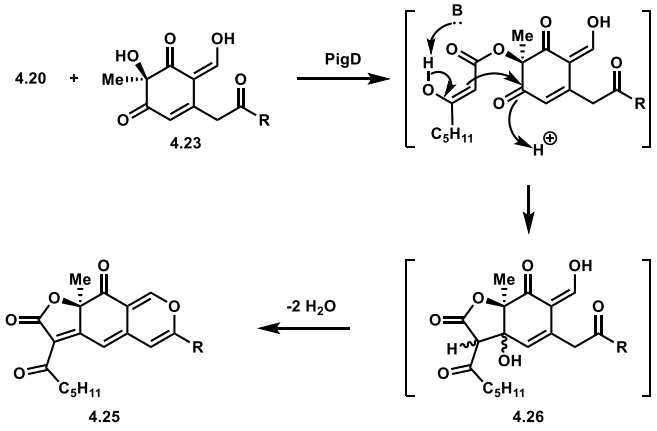


Figure 4.5. Mechanistic hypothesis for the observed condensation regioselectivity afforded through acylation of dearomatized substrates by PigD.

situ using AzaH. We thus hypothesize that direct acylation of this dieneol form could confer the observed selectivity in a substrate-dependent manner (Figure 4.4).[34]

4.4 Conclusion and Future Directions

In this Chapter, we described the total synthesis of an azaphilone natural product and several analogs through a one-pot, two-enzyme sequence. Two ATs from azaphilone biosynthetic pathways were initially investigated for their ability to catalyze an acyl transfer through clarified cell lysate reactions with an azaphilone bicycle substrate. Of these two enzymes, PigD displayed signifcant acylation activity, determined by UPLC-DAD and LC-MS. After moving to a purified enzyme platform, we demonstrated the ability of PigD to not only perform an acylation reaction on multiple dearomatized substrates generated *in situ*, but also that these reaction conditions afford linear tricycles exclusively. This unprecedented cyclization selectivity provided a means to complete the first reported total synthesis of the natural product rubropunctatin. Preperative-scale reactions with two other substrates under these conditions afforded synthetic analogs of the natural product, as well. NMR analysis of these analogs confirms they are also linear tricycles, suggesting that this unexpected cyclization selectivity holds true for multiple substrates.

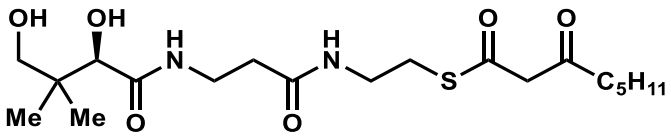
Moving forward, we will continue exploring the substrate scope of PigD with other thioester substrates and dearomatized intermediates in the system described above towards constructing a diverse library of linear tricyclic azaphilones. We intend to also further probe the source of the observed cyclization regioselectivity through generation of the uncylized, acylated intermediate (see **4.15**) synthesized through conventional chemical means. Due to the ease in which the reactions can be conducted in a 96 well

plates, we also have plans to collaborate with the Sexton Lab at the University of Michigan to screen compounds synthesized with this methodology for biological activity in their high-throughput and high-content assays in search of novel pharmacophores.

4.5 Experimental

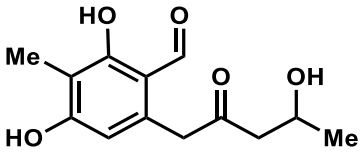
I. Substrate synthesis

All reagents were used as received unless otherwise noted. Reactions were carried out under a nitrogen atmosphere using standard Schlenck techniques unless otherwise noted. Solvents were degassed and dried over aluminum columns on an MBraun solvent system (Innovative Technology, inc., Model PS-00-3). Reactions were monitored by thin layer chromatography using Millipore 60 F_{254} precoated silica TLC plates (0.25 mm) that were visualized using UV, *p*-anisaldehyde, CAM, DNP, or bromocresol stain. Flash column chromatography was performed using Machery-Nagel 60 μm (230-400 mesh) silica gel. All compounds purified by column chromatography were sufficiently pure for use in further experiments unless otherwise indicated. ^{1}H and ^{13}C NMR spectra were obtained in $CDCl_3$ at rt (25 °C), unless otherwise noted, on Varian 400 MHz or Varian 600 MHz spectrometers. Chemical shifts of ^{1}H NMR spectra were recorded in parts per million (ppm) on the δ scale. High resolution electrospray mass spectra were obtained on an Agilent HPLC-TOF at the University of Michigan Life Sciences Institute.



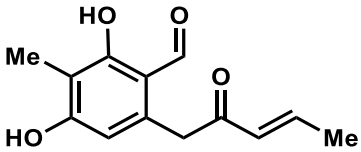
(*R*)-S-(2-(3-(2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl)3-oxooctanethioate (4.20)

Prepared as previously by Chen and coworkers.[25] **^{1}H NMR** (600 MHz, CDCl3) δ 7.52 – 7.47 (m, 1H), 6.95 (dt, *J* = 11.0, 5.7 Hz, 1H), 3.95 (s, 1H), 3.67 (s, 2H), 3.54 – 3.32 (m, 7H), 3.07 – 2.98 (m, 2H), 2.49 (t, *J* = 7.4 Hz, 2H), 2.40 (t, *J* = 6.2 Hz, 2H), 2.12 (t, *J* = 7.7 Hz, 1H), 1.54 (p, *J* = 7.4 Hz, 2H), 1.29 – 1.21 (m, 4H), 0.93 (d, *J* = 2.9 Hz, 3H), 0.87 (s, 3H), 0.85 (t, *J* = 7.1 Hz, 3H). All spectral data matches the published values.[25]



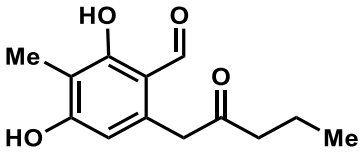
2,4-Dihydroxy-6-(4-hydroxy-2-oxopentyl)-3-methylbenzaldehyde (3.S12)

Prepared as previously reported by Baker Dockrey *et al*.[43] (see Chapter 3 for spectrum)



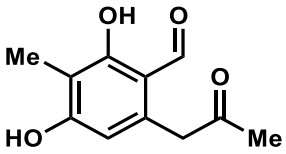
(*E*)-2,4-Dihydroxy-3-methyl-6-(2-oxopent-3-en-1-yl)benzaldehyde (3.4)

Prepared as previously by Pyser *et al*.[35] (see Chapter 3 for spectrum)



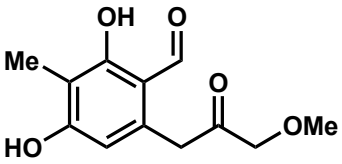
2,4-Dihydroxy-3-methyl-6-(2-oxopentyl)benzaldehyde (3.S9)

Prepared as previously reported by Baker Dockrey *et al*.[43] (see Chapter 3 for spectrum)



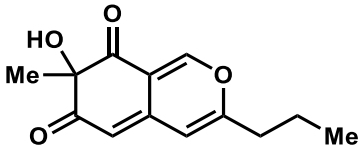
2,4-dihydroxy-3-methyl-6-(2-oxopropyl)benzaldehyde (3.7)

Prepared as previously reported by Baker Dockrey *et al*.[43] (see Chapter 3 for spectrum)



2,4-dihydroxy-6-(3-methoxy-2-oxopropyl)-3-methylbenzaldehyde (3.S3)

Prepared as previously by Pyser *et al*.[35] (see Chapter 3 for spectrum)



7-hydroxy-7-methyl-3-propyl-6H-isochromene-6,8(7H)-dione (4.24A)

Prepared as previously reported by Baker Dockrey *et al*.[2]

II. Plasmids and Proteins

Plasmids: The plasmid encoding *pigD* (AGI63864.1),[25] *cazE* (XM_001225292),[32] *tran1* (from *Stachybotrys* chartarum, strain CBS 109288 / IBT 7711, A0A084B9Z3), *tran2* (from *Pseudogymnoascus sp. VKM F-4513*, A0A094B1C4), *tran3* (from *Pseudogymnoascus sp. VKM F-103*, A0A094EQW5), *tran4* (from *Pseudogymnoascus sp. VKM F-4520*, A0A094J5P9), *tran5* (from *Exophiala spinifera*, A0A0D1YCS0), *tran6* (from *Aspergillus parasiticus*, strain ATCC 56775 / NRRL 5862 / SRRC 143 / SU-1, A0A0F0IK60), *tran7* (from unknown organism, A0A0F8UU93), *tran8* (from *Aspergillus rambellii*, A0A117E0B7), *tran9* (from *Penicillium patulum*, A0A135LJ73), *tran10* (from *Didymella*

rabiei, A0A163CB57), *tran11* (from *Valsa mali var. pyri.*, A0A194UPD1), and *tran12* (from *Valsa mali.*, A0A194VYW6) were synthesized by Twist Biosciences and cloned into a pET-28a(+) vector.

Non-optimized *pigD* Sequence

ATGGAGGATCCTGCTCGGACGCAAGACATCCTTGGACAGCTGCCAATCCTCAAGG
CATACAACCATATTCTCCTGGGGTTTGCATTGTCCGAGGATATCTCCCGGGAATCC
GTCGTCCAAGCGCTCAACGCAGCAGCCCTGCAGCTAGCAACGTCCGTCCCATGGA
TCGGAGGCAAGGTCGTGAACGTGGGAAGCGGACCCGGCAACACTGGGCTATTCA
GGAGCGTGCCATGCGAGCTGTTCGCGCCTCCGAATTCGATCCTTCGAGTCAAGGA
TGTGACGGCGGACTATCCCTCTTACGAGGAGATCGTGAGCGCCAAAGGCCCCATC
TCGATGCTGGACGGCAGCATCATTGCACCCAAGCCTGCATTTCCTGTGAGCTATGT
AGACTGTGAATCCGATCCAGCCCCGGCCCTCCTCATCCAGGCCACTTTCCTCAAG
GGTGGTGTGCTGTTGGACTTCGCCGCGCAGCACAATCTGTCCGACGGCGGCGGC
GTCATCCAGATGATAAACCTGGTCGCTACCACCCTGCGCGGCGAGAAGATCCCCG
AAAAGGCCATAGTACAGGCCAATCGAGACCGCAGAGATGTCATCCGGCTTCTGGA
TCCCGCAGAACCCATGCTGGACCACAGCCATCTCGTCCGTCCGCCGCCGTCCGC
GATCCCCGCCAACCCCGTCGTCTCGCCCGATAACTTCATCTGGCAGTATTTTCGCT
TTTCCGCGTCGACGCTGGGTGCGTTGAAGAACATCGCTTCGAATCCGGCCGACTT
CGACCCGTCTGTCAAGTTTATCTCGACGGATGACGCCCTGTGCGCGTTCTTATGG
CAGCGCATCGCGACCGTCCGTCTACGCCGCCGCCAGACGCCGGATGACCTGTGC
AAGATGACACGGGCCGTCGACATCCGCCGTACTCTCCAGGTGCCATCGGAGTACA
TGGGCGTGATGGTATACAATGTGAGCGGTCGACTGCCTCTCGGCCAACTGGCGAC
GGCTTCGCTGGCCCGCGCGGCGTCCGAGTTGCGCAAAGCCCTAAACAGCATCGA
CGAGTATGCAGTGCGCAGCTTCGCCACATTTGTAGCGCGCCAGCCGGATAAGTCC
ACCCTGGCGTACGCGGGGAAGTTCAATCCTGATGTTGACATGGGGGTATCCTCTA
TGGCGTCGGTGCCGCTGTATCGCGCCGATTTTGGCCCACTGGGCGCTCCGGGCC
TCGTGCGACGGCCCAATTTTGCACCGGTCCTGAGCACCATCTATGTGATGCCGCA
GACTGTGGAAGGGGACGTTGATGTCTTGATCTGTCTGACGAGGGAGGATATTGGG
GCCTTGCGGGCGGACCCCGAATGGACTGCGTATGCGGAGTATATAGGGTGATTTA
CACCATCGAAGAGAGGCAG

PigD Protein Sequence

MEDPARTQDILGQLPILKAYNHILLGFALSEDISRESVVQALNAAALQLATSVPWIGGKV
VNVGSGPGNTGLFRSVPCELFAPPNSILRVKDVTADYPSYEEIVSAKGPISMLDGSIIAP
KPAFPVSYVDCESDPAPALLIQATFLKGGVLLDFAAQHNLSDGGGVIQMINLVATTLRG
EKIPEKAIVQANRDRRDVIRLLDPAEPMLDHSHLVRPPPSAIPANPVVSPDNFIWQYFR
FSASTLGALKNIASNPADFDPSVKFISTDDALCAFLWQRIATVRLRRRQTPDDLCKMTR
AVDIRRTLQVPSEYMGVMVYNVSGRLPLGQLATASLARAASELRKALNSIDEYAVRSF
ATFVARQPDKSTLAYAGKFNPDVDMGVSSMASVPLYRADFGPLGAPGLVRRPNFAPV
LSTIYVMPQTVEGDVDVLICLTREDIGALRADPEWTAYAEYIG

Non-optimized *cazE* sequence

ATGGGTTCCAACCAGGACGACGTGATCACGCAGCAGCCCGGCGCCGGAGAGGAT
GGCGTGGGCAGCAGCAACGTTAACCTCGACCGACTTCCCGACCTCATGGGCCAG
TTCCCCGTTCTTAACGCTTACACTCACATAACCTTTGGATTTGAACTCCCACTCGAT
GTCGACCGCGACGCTATTACTTCCGCCCTTCAGACCAGCCTCGATCGGCTGATAG
AGCTGATCCCGTGGCTGGGATGGCAAATCGGTTTGGAATCAGGGGTACGCACGAC
GGTTCCCTGGCCCGAGGATACGCCCAGGGTCCTACTGCACGTCAAAAACTGCGAC
GACATTGTTGTACCCATGGCGCAGCTTCTTGCCGCAGGCATTCCCATCAACAAACT
TGATGCCAAAGTGTTGGCCCCCTGGCCGGCACTCCCGCAACCCCACGGGCTCAC
AGGTCCCGTACCAGTCATCACCCTCCAGGCCAACTTTGTCCGCGGCGGGCTCATC
CTCAACTTCAGCTCGCATCATACGGTGATGGACGGCACGGCCAACCTGCAAGTCC
CGAAGCTATTCGCCGCCATCCTTAACGGCGATACTATCCCAGAAGCCGACCTCCA
GCAGGCGAACCGCGACCGACGTGGCCTGATCGAGCTCATCCCGCGCAGCGAGCC
TATGAAGGACTACTCCCACTTGCGAGGACCGCCAGGATATCGATTCACGCTCCCT
GCCAGCGCGCCGATATGGTGCCACTTCAACATGCCCATGGCCTCCCTCTCCAAAC
TCACCAAGTCAGCGCGCGACCCCTCCCGACCCGTCAACGAAGACGACGTTCTCAT
CGCCTTCTTCTGGCAGCGCCTCTGCGCCGCCCGCATCGCCCGCGGCATCGCTCG
GGACACGGTGTCCAAGATCTCGCGGGCCATCGACGCGCGCGTGGCGCTGGGGAT
CCCGCCAACCTACCTCGGCGCGCAGGTGCACATGGCCATCACGCGGCTGCCCAT
GGGCCAGGTGGCGTCGCTCACGCTGCTGCAGGTGGCGCAGGTGCTGCGGCGGG
CGCTGGCCGACGCCAACACGCCGTGGGCGATCCGCAGCCTGGTGACCTTCATCG
CGCGCGAGCCAGACCGCTCCCGCCTGCTGTACAACGGCCCCCACGACGGCAACA
CCGACGTCGGCGCCACGTCGAGCCTGAACATGGCCCACGCCCGCCCGCCGCATT
GGGGACCGGTGCTCGGCCCGTCCCGCTTCTTTAGTCGCTCGAATGGGGCGCCTA
TTCCGGGGTCGTTTACAGTCTATTCGCCGGACGGCGCCGGTTTTGTGCGCATGTG
TATCTGCTTGCCGGTGGATGATCTGGAGGCGTTGAAGAAGGATGCGTTGTGGAAG
CAGCATACGAAGTTTGTTGGGTGA

CazE Protein Sequence

MGSNQDDVITQQPGAGEDGVGSSNVNLDRLPDLMGQFPVLNAYTHITFGFELPLDVD
RDAITSALQTSLDRLIELIPWLGWQIGLESGVRTTVPWPEDTPRVLLHVKNCDDIVVPM
AQLLAAGIPINKLDAKVLAPWPALPQPHGLTGPVPVITLQANFVRGGLILNFSSHHTVM
DGTANLQVPKLFAAILNGDTIPEADLQQANRDRRGLIELIPRSEPMKDYSHLRGPPGYR
FTLPASAPIWCHFNMPMASLSKLTKSARDPSRPVNEDDVLIAFFWQRLCAARIARGIAR
DTVSKISRAIDARVALGIPPTYLGAQVHMAITRLPMGQVASLTLLQVAQVLRRALADANT
PWAIRSLVTFIAREPDRSRLLYNGPHDGNTDVGATSSLNMAHARPPHWGPVLGPSRF
FSRSNGAPIPGSFTVYSPDGAGFVRMCICLPVDDLEALKKDALWKQHTKFVG

Codon optimized *tran1* sequence

ATGAGCACCATGGCGAAAAGCCCGGAAGCGAACAACCTGCATCAGGATGTGATTG
CGCAGTTTCCGATTCTGAACGGCTATACCCATACCGTGGGCGCGTTTAGCCAGCC
GCTGAACGTGAGCCGCCTGTTTATTATTGATGAAATTCAGACCGCGTATGATGAAC
TGCGCGTGCAGATTCCGTGGCTGGCGCATCAGGTGGTGGTGGTGGATGCGGGCC
CGGGCAAAAGCGGCTATATTACCACCGCGCCGTGGCCGAGCAGCGCGCCGCCGA
ACGATGTGACCTATGAAGAAAAAGATGATGCGTTTCCGAGCCTGAACACCCTGATT
AAAAGCGGCGGCAGCTTTCTGGCGACCAAAGATCTGGTGGGCTATCCGGGCCTG

CCGGAACCGCATGGCCTGCATCCGACCCCGGTGGCGACCATTCGCCTGGTGTTTA
TTACCGGCGGCGTGCTGGTGGTGCTGAGCACCCATCATAACATTGTGGATGGCAT
TGGCCTGATGCAGATGTGGGATTATCTGGATATTCTGATGGGCGGCGGCGCGATT
AGCCGCCAGGATGCGCGCAGCGCGAACGCGGATCGCGCGCGCGTGCTGCCGCT
GATTGCGCCGGGCGAACCGGTGAAAGATTATAGCCATCTGATTCGCCCGAACCCG
TGGCCGCTGCCGCCGCCGCCGAAAACCGAATGGCGCCTGTTTAAAATGCATCCGT
GGGCGCTGGCGGAAATTCGCAGCCGCGCGCGCGATGGCACCGATCAGCGCGCG
AGCGCGCGCCCGGCGAGCAGCGATGATGCGCTGACCGCGTTTTGCTGGCAGCGC
GTGAGCGCGATGCGCCTGGCGAGCGGCCGCGTGACCGGCGATCAGGTGAGCAA
ATTTGGCCGCGCGGTGAACGGCCGCAGCGCGATGGGCCTGGATAGCAGCTATCT
GTTTCATATGATGCTGCATACCGAAACCCGCCTGCCGATTGAACAGATTGCGCGCA
GCACCCTGGCGGAACTGAGCACCCAGCTGCGCAAAGATCTGGATGCGGCGCGCA
CCGAATGGAGCGTGCGCAGCTATGCGACCTTTCTGGCGGGCGTGGCGGATAAAA
CCCGCCTGCTGTATGGCGGCATTACCAACCCGCAGACCGATCTGGGCGGCACCA
GCACCATGCATTGGGCGAGCCGCCGCCCGATTCGCCTGGGCCTGCTGGGCGATT
GCCATCTGATTCGCAAACCGGAAGGCATGCCGCTGCCGGGCTGCCTGTATTTTAT
GCCGAGCGGCGGCACCAGCGGCGTGGTGCAGCTGCTGCTGTGCCTGCCGAAAG
AAGAACTGGATGCGCTGCAGGAAGATGCGGAATGGAAACATTATACCGAAAGCGG
CGGCCGCCGCGTGGATGGCCCGCGCCTG

Tran1 Protein Sequence
MSTMAKSPEANNLHQDVIAQFPILNGYTHTVGAFSQPLNVSRLFIIDEIQTAYDELRVQI
PWLAHQVVVVDAGPGKSGYITTAPWPSSAPPNDVTYEEKDDAFPSLNTLIKSGGSFLA
TKDLVGYPGLPEPHGLHPTPVATIRLVFITGGVLVVLSTHHNIVDGIGLMQMWDYLDILM
GGGAISRQDARSANADRARVLPLIAPGEPVKDYSHLIRPNPWPLPPPPKTEWRLFKMH
PWALAEIRSRARDGTDQRASARPASSDDALTAFCWQRVSAMRLASGRVTGDQVSKF
GRAVNGRSAMGLDSSYLFHMMLHTETRLPIEQIARSTLAELSTQLRKDLDAARTEWSV
RSYATFLAGVADKTRLLYGGITNPQTDLGGTSTMHWASRRPIRLGLLGDCHLIRKPEG
MPLPGCLYFMPSGGTSGVVQLLLCLPKEELDALQEDAEWKHYTESGGRRVDGPRL

Codon optimized *tran2* sequence
ATGGATTTTCGCCCGTATGTGGAAGATACCAGCGCGAGCGCGCCGGAACAGCCG
GCGCAGGCGGAAAGCAAAACCCAGGTGGATAGCATTGATGAACTGCATCAGGATG
TGGTGAGCCAGTTTCCGTTTCTGAACGGCTATAGCCATATTGTGTTTCTGTTTCAGC
CGGATGCGGATACCAGCCGCGAAACCATTATTGCGACCGTGCATACCGCGCTGGC
GAAAGTGACCGCGCAGGTGCCGTGGCTGGCGGGCCAGGTGCTGCATAACCCGG
GCCCGCCGGGCAACAGCGGCCATTATATGCTGGCGCCGTGGCCGCTGAACGCGC
ATAAAAACGATATGGTGCGCGTGGAAGATTGCGAAGATCGCATGCCGCCGATTGC
GCAGATTGTGCAGGCGGGCGCGCCGATTAGCATGCTGGATGCGAAAATTCTGACC
CCGTTTCCGAGCCTGCCGCTGCCGCATGGCCTGGAACCGCCGCTGCCGGTGATG
CTGCTGCAGGTGAACTTTCTGCGCGGCGGCGTGATTATTAACCTGAGCACCCATC
ATATGGCGGTGGATGGCAACGGCATTGTGCAGATTATGCGCCTGCTGGCGACCGT
GCTGCAGGGCCAGGAAATTGCGGCGGCGGATCGCGAACAGGCGCGCCGCGATC
GCCGCCGCGTGGTGCCGCTGATTCCGCGCGGCGAACCGGTGAAAGATCATAGCC
ATCTGCGCGCGCCGCCGGGCCATGTGACCGTGCCGCCGAGCAGCCCGCCGAAAT
GGTGCTATTTTAAACTGCCGGTGAGCGCGCTGCCGACCCTGAAAAAACTGGCGAG

CCCGCCGAGCCAGCAGCAGCAGCTGAGCGAAAACGATGTGATTTGCGCGTTTTGC
TGGCAGCGCATTACCGCGATGCGCCTGGCGCGCGGCTTTAGCCCGGATACCATG
ACCAAAATGACCCGCATTATTGATGCGCGCGCGGCGCTGGGCGTGCCGATTACCT
ATATGGGCCATCTGACCTATTATGCGGTGGCGCAGCTGCCGATGCATCTGGTGAC
CAGCCTGCCGCTGAGCAGCATTGCGCAGACCCTGCGCCGCGAACTGAACGCGGC
GAACACCACCTGGGCGATTCGCAGCTATGCGACCTTTCTGGCGCGCGAACCGGAT
AAAAGCCGCCTGGCGTATGCGGGCCTGCGCAACCTGGATACCGATCTGGGCACC
ACCGCGTTTACCGCGAGCCAGAGCACCAGCCATGATGATGGCAGCGCGATTCCG
GAAGATTTTGGCCCGGTGCTGGGCCGCCTGAAATATACCCGCCGCCCGAACTGCA
GCCCGCTGGTGGGCGGCATTACCATTTGCCCGGTGGAAAGCGGCGCGATTCCGA
TTGTGATGTGCCTGCCGGGCGCGGATCTGGAAGGCCTGGAAAAAGATCGCGAAT
GGCGCCGCTATGCGCGCTATATTGGC

Tran2 Protein Sequence
MDFRPYVEDTSASAPEQPAQAESKTQVDSIDELHQDVVSQFPFLNGYSHIVFLFQPDA
DTSRETIIATVHTALAKVTAQVPWLAGQVLHNPGPPGNSGHYMLAPWPLNAHKNDMV
RVEDCEDRMPPIAQIVQAGAPISMLDAKILTPFPSLPLPHGLEPPLPVMLLQVNFLRGG
VIINLSTHHMAVDGNGIVQIMRLLATVLQGQEIAAADREQARRDRRRVVPLIPRGEPVK
DHSHLRAPPGHVTVPPSSPPKWCYFKLPVSALPTLKKLASPPSQQQQLSENDVICAFC
WQRITAMRLARGFSPDTMTKMTRIIDARAALGVPITYMGHLTYYAVAQLPMHLVTSLPL
SSIAQTLRRELNAANTTWAIRSYATFLAREPDKSRLAYAGLRNLDTDLGTTAFTASQST
SHDDGSAIPEDFGPVLGRLKYTRRPNCSPLVGGITICPVESGAIPIVMCLPGADLEGLEK
DREWRRYARYIG

Codon optimized *tran3* sequence
ATGGGCAGCATTGCGGAAGATTATAACAACTTTGCGCGCTATCAGGATATTTTTGG
CCAGCTGCGCCTGCTGAAAAGCTATACCCATTTTCTGCTGTGCTTTCCGATTCCGG
AAGGCACCAGCCATGATACCATTGTGGAAGGCCTGCGCAGCGCGACCCTGAAAGT
GACCAGCACCTTTCCGTGGCTGACCGGCAAAGTGATTAACGAAGGCAGCGGCAAA
GGCAGCAGCGGCCTGTTTAAAGTGGCGCATTGCCCGCAGTGGGAACCGCCGAAC
AGCATTCTGCGCGTGAAAGATTGCAGCCATATTTGCCCGAGCTTTGCGGAAATTAT
GCGCGAAAAAGGCCCGATTAAATTTTTTAACCCGAAAGATCTGGCGCCGACCGTG
GCGTTTCCGCAGAGCTATCAGGAAACCGATGATGATCCGGCGTGCGTGTTTGCGC
TGCAGGCGAACATTGTGGAAGGCGGCCTGCTGCTGGATATTGCGGCGCAGCATAA
CATTATGCCGGGCGGCGGCATTCTGCAGTTTCTGAGCCTGCTGAGCAAAGTGATG
AACGGCGGCGAAATTACCACCTTTGAAATTGAACAGGGCAACCGCGATCGCCGCA
ACATGGTGAAACTGCTGGGCCCGGATGAACCGATTATTGATCTGGGCCGCTATCG
CCGCCCGAGCCTGCTGGAAGCGAGCATTCCGGCGGTGCCGGCGCCGCATGGCA
AATGGTGCATGTTTCGCTTTAGCGCGGCGAGCATTAACGCGCTGAAAAAAATGGC
GAGCGATACCAGCAAATTTGTGGAAGGCATTACCTTTGTGAGCAGCAACGATACCC
TGACCAGCTTTATTTGGAAACGCATTGCGGCGATTCGCCTGCGCCGCATGGTGGA
TCCGAACGCGAGCAGCAAATTTGTGCGCGCGGTGGATGTGCGCCCGACCATGGG
CGTGCCGAACGAATATATGGGCCATATGGTGTATAACACCTTTACCACCCTGACCA
TGCGCGAAGTGGATGAACTGCCGTTTCCGACCCTGGTGAGCCTGATGCGCAAAAA
CCTGCAGGAAGATGTGAACGAATATGCGGTGCGCAGCTTTGTGACCCTGCTGGAT
CGCACCCCGGATAAAACCACCATTATGTATGGCGGCGAAATGCAGCCGGATACCG

ATGTGGCGTTTACCAGCCTGGCGCAGAGCGATCTGTATAACGTGAGCTTTGGCGC
GCTGGGCAAACCGGCGCTGCTGCGCCGCCCGAACTTTATTCCGCGCACCACCAC
CAGCATGATTTTTCCGAAAGCGCCGGATGGCAGCCTGGATGTGATGGTGTGCCTG
AGCGAAGATGATCTGGATCAGCTGAAAGCGGATCCGGTGTGGAGCAGCCATACCG
AACTGATTGATAAAGAT

Tran3 Protein Sequence
MGSIAEDYNNFARYQDIFGQLRLLKSYTHFLLCFPIPEGTSHDTIVEGLRSATLKVTSTF
PWLTGKVINEGSGKGSSGLFKVAHCPQWEPPNSILRVKDCSHICPSFAEIMREKGPIKF
FNPKDLAPTVAFPQSYQETDDDPACVFALQANIVEGGLLLDIAAQHNIMPGGGILQFLS
LLSKVMNGGEITTFEIEQGNRDRRNMVKLLGPDEPIIDLGRYRRPSLLEASIPAVPAPHG
KWCMFRFSAASINALKKMASDTSKFVEGITFVSSNDTLTSFIWKRIAAIRLRRMVDPNA
SSKFVRAVDVRPTMGVPNEYMGHMVYNTFTTLTMREVDELPFPTLVSLMRKNLQEDV
NEYAVRSFVTLLDRTPDKTTIMYGGEMQPDTDVAFTSLAQSDLYNVSFGALGKPALLR
RPNFIPRTTTSMIFPKAPDGSLDVMVCLSEDDLDQLKADPVWSSHTELIDKD

Codon optimized *tran4* sequence
ATGGCGGAAGAAGCGCTGCTGCAGGATGTGATGGGCCAGTTTCCGCAGCTGAAAA
CCTATAACCATGGCAGCCTGATTTTTAGCATGCCGGATGGCGTGAGCCGCGAAAG
CCTGGTGACCGCGCTGGAAGCGAGCACCGGCAAAATTCTGAGCGCGATTCCGTG
GCTGCATGAACAGGTGGTGCATGTGGGCCAGGCGCCGGGCACCAGCGGCAAATT
TGAACTGGCGCCGTGGCCGGCGGATGCGCCGAAAAACACCCTGATTCGCGTGCG
CGATTGCAGCGATCTGCTGCCGAGCCATGAAGAAATGCTGAAAAGCCAGGGCCCG
GCGAGCATGCTGGATGGCAACCTGATTTGCGCGGTGCCGGGCTTTCCGCTGCGC
TATGATGAAGCGGTGATTGGCCCGGCGCCGGCGGCGATTATTCAGGTGAACTGGA
TTAAAGGCGGCGTGATTCTGACCTTTAGCAACCAGCATAACGTGATGGATGGCAG
CGGCGTGTTTCAGCTGATTGCGCTGCTGAGCACCGCGATGAACGGCGAAGAAATT
AGCAAAAGCGCGATTGAAGAAGGCAACCGCGATCGCAAAACCGTGGTGCCGCTGT
ATGGCCCGGATGAACCGATTCGCGATCATAGCCATATGCTGATTAAAACCCCGCT
GAGCGCGGCGGCGGCGCCGAACAAAGCGCAGAGCCTGAGCAGCTGGGTGTTTCT
GCGCTTTTGGAAAAAAAGCGTGCCGGAAATTAAAGCGATGGCGACCGATCCGGTG
GGCTATGATAAAAGCGTGCCGTTTATTACCAGCAACGATGCGATTAGCGCGTTTTA
TTGGAAACGCCTGGCGATTGCGCGCGTGCGCAACGGCCAGGATGGCCAGGCGCT
GAGCAAATTTAGCCGCGCGATTGATGCGCGCGCGGTGATGGGCGTGAGCCGCGA
ATATATGGGCCAGATGGTGTATTTTAGCGCGACCTGGCTGACCTATCAGGAACTG
GTGGATCTGCCGCTGAGCACCATTGCGAGCCGCATGCGCAAAAGCATGAACGAAA
GCAACAACGAATTTAGCGTGCGCAGCTATGCGACCTTTGTGAGCCGCATTCAGGA
TCGCACCACCATTGCGTATGGCGGCCCGTTTAACCGCGAAACCGATATTGCGAGC
AGCAGCATGGCGCAGGCGGCGGTGGTGCTGAAATTTGGCATTCTGGGCACCCCG
GAACTGGTGCGCCGCCCGAACCTGGCGCCGATTCCGGGCACCCTGTATTTTTTTC
CGCCGGATGGCAGCGGCGATCTGAACCTGCTGATGTGCCTGACCCAGCAGGAAC
TGGAAGCGATGCGCAGCGATCCGGAATGGAGCCAGTGCACCCAGTATATTGGC

Tran4 Protein Sequence
MAEEALLQDVMGQFPQLKTYNHGSLIFSMPDGVSRESLVTALEASTGKILSAIPWLHE
QVVHVGQAPGTSGKFELAPWPADAPKNTLIRVRDCSDLLPSHEEMLKSQGPASMLDG

NLICAVPGFPLRYDEAVIGPAPAAIIQVNWIKGGVILTFSNQHNVMDGSGVFQLIALLSTA
MNGEEISKSAIEEGNRDRKTVVPLYGPDEPIRDHSHMLIKTPLSAAAAPNKAQSLSSWV
FLRFWKKSVPEIKAMATDPVGYDKSVPFITSNDAISAFYWKRLAIARVRNGQDGQALSK
FSRAIDARAVMGVSREYMGQMVYFSATWLTYQELVDLPLSTIASRMRKSMNESNNEF
SVRSYATFVSRIQDRTTIAYGGPFNRETDIASSSMAQAAVVLKFGILGTPELVRRPNLAP
IPGTLYFFPPDGSGDLNLLMCLTQQELEAMRSDPEWSQCTQYIG

Codon optimized *tran5* sequence

ATGACCGCGAAACTGCGCCAGCCGGGCATTTTTGGCCAGCTGTGCTGGGATACCT
ATACCCTGATTCTGCTGGGCTTTCCGGATCGCCAGGGCAGCAGCCGCGATCAGGT
GCTGAGCACCCTGGATAGCGCGGCGCTGCGCCTGCTGAGCGCGTATCCGAGCCT
GGCGGGCCAGGTGGTGAACCGCGGCCGCACCGCGACCAACAGCGGCACCTATG
AAATTGTGCCGTATCCGCCGCATGAAAACAAAAGCCCGGTGCGCCGCAAAGATTG
CAGCGATCTGTGCCCGAGCTATGATGAAATTCTGAAAGCGGATGCGCCGTTTTTTA
TGCTGGATGGCGATATTCTGTGCCCGATGAAAGGCATGGGCTATGCGTATGGCCC
GGAAACCGAATTTCCGGTGTTTATTGTGCAGGCGAACTTTGTGAAAGGCGGCCTG
CTGCTGTGCTTTGCGTGCATGCATAACGCGCTGGATATGAACGGCCAGGGCACCG
TGCTGAAAATGTTTGCGGCGGCGGGCCGCGGCGATGAACTGGATCCGACCCTGC
TGGCGGCGGAAAACCTGGATGCGGATAGCATTGTGCCGCTGCTGAAACCGGGCG
AAGCGAGCGCGAGCCATGATGTGATGCGCCGCCCGAGCACCCTGACCCCGAGCC
AGACCACCGCGGGCCCGCCGCGCGCGCTGCCGTGGAACTATTGGCGCTTTACCG
GCGAAGGCCTGGCGCAGCTGAAAAAAACCGCGAGCAGCGGCAGCGAATGGGTGA
GCACCAACGATGCGCTGACCGCGTTTCTGGTGCAGCGCCTGACCGCGGTGCGCG
TGGCGGCGGGCCGCGTGGCGCCGGATGAAGAAGTGCATCTGTATCGCGCGGTG
GATAGCCGCAGCATTCTGAAACCGCCGGTGCCGGATGGCTATCTGGGCCATCTGG
TGAGCCTGGCGGATACCAAATGGAGCACCGCGCAGGATATTTGCGAAAGCAGCTT
TGCGGATCTGGCGGTGAAAGTGCGCAAAAGCCTGCGCGAAGTGGATGATCATTTT
GTGCGCAGCCTGGCGACCCTGATTAAAAGCACCGAAGATAAAAGCACCATTTTTTA
TGGCGCGAAAGCGAAACCGGGCCGCGATTTTATTTGCAGCAGCTGGGCGCAGCT
GCATTGGCTGAGCAACTGCGATTTTGGCCCGGGCCTGGGCACCGTGGATTTTGTG
CGCCGCGCGCGCCTGCCGGGCGTGCCGGATCTGACCTATATTATGCCGAAAAAC
CATAAAGGCGATATGCATATTGCGGCGAGCACCTTTATGGATGATTTTGTGGGCCT
GGTGAACGATCAGCAGTGGCGCGAATTTGCGCAGCTGGTGGGC

Tran5 Protein Sequence

MTAKLRQPGIFGQLCWDTYTLILLGFPDRQGSSRDQVLSTLDSAALRLLSAYPSLAGQ
VVNRGRTATNSGTYEIVPYPPHENKSPVRRKDCSDLCPSYDEILKADAPFFMLDGDILC
PMKGMGYAYGPETEFPVFIVQANFVKGGLLLCFACMHNALDMNGQGTVLKMFAAAG
RGDELDPTLLAAENLDADSIVPLLKPGEASASHDVMRRPSTLTPSQTTAGPPRALPWN
YWRFTGEGLAQLKKTASSGSEWVSTNDALTAFLVQRLTAVRVAAGRVAPDEEVHLYR
AVDSRSILKPPVPDGYLGHLVSLADTKWSTAQDICESSFADLAVKVRKSLREVDDHFV
RSLATLIKSTEDKSTIFYGAKAKPGRDFICSSWAQLHWLSNCDFGPGLGTVDFVRRAR
LPGVPDLTYIMPKNHKGDMHIAASTFMDDFVGLVNDQQWREFAQLVG

Codon optimized *tran6* sequence

ATGGGCATTTTTGAACATGTGCAGGATGTGATTGGCCAGCTGCCGGTGCTGAAAA
GCTATAGCCATATGCTGATTTGCTTTCCGGTGCAGGATGATAAACGCGAAGTGGC
GATTCAGGAACTGGAACGCGCGGTGCGCCTGGTGATGAAAACCTTTCCGTATCTG
AGCGGCAAAGTGATTAACGAAGGCAGCGGCGCGGGCAGCAGCGGCACCTTTAAA
GTGACCAGCTATAAAGAATGGGAAAGCGAAACCCATGTGTTTGTGCGCGTGCAGG
ATCGCACCACCGAATGCCCGGCGTATGATGAACTGTGCGCGGCGCGCGGCCCGA
GCAGCATGCTGCCGGGCCATCTGCTGAGCAGCCGCGTGGCGTTTCCGGAAACCT
ATCAGGATATTGAAGAAGATCCGGCGCCGGTGCTGGATTTTCAGGCGAACATTGT
GCGCGGCGGCCTGCTGCTGGATCTGGCGGCGCAGCATAACATTATTGATGGCAC
CGGCATTTTTCAGATTATTAACCTGCTGGCGACCGCGCTGCGCGGCGATCAGTTTC
CGCTGTTTCAGCTGCATGAAGGCAACCGCGATCGCCGCGATCTGATTCGCCTGCT
GGGCCCGGATGAAGCGCTGCTGGATCATAGCGAACTGAAACCGCCGGTGATTATG
AAAGCGCCGCCGCCGAGCGATATTCTGGCGCCGTATAAATGGCGCTATTATCGCT
TTCCGGTGGATAGCGTGAACAAAATTCGCGATCTGGCGAACAGCAAACCGGAAGA
TTTTGATCCGAGCACCGAAAGCCTGAGCCTGAACGATGCGATTACCGCGTTTTGCT
GGCAGCGCATTACCGCGATTCGCCTGCGCAAACTGAAAACCCCGACCGCGTTTAG
CAAACTGAGCCGCGCGGTGGATTTTCGCCGCATTATGCGCCTGACCCCGGCGTAT
CTGGGCCATATGGTGCGCGTGTGCAACACCCGCCTGACCTTTGAAGATATTGTGG
AAAGCAGCCTGAGCCGCCTGGCGAGCCTGCTGCGCAAAGATGTGCAGGAAATTA
GCAACGAATATGCGCTGCGCAGCTATGTGACCTTTATTGCGAACGAACCGGATAAA
AGCGATATTGCGTATGGCGGCAGCTTTAACGCGCAGACCGATTTTAGCTGCAGCA
GCATTGCGCATGTGAAAGCGCCGGATTTTGGCCCGCTGGGCAAACCGGGCCTGA
TGCGCCGCCCGACCTTTCAGCCGCTGCCGTGCAGCAGCTATATTGCGCCGATGCT
GCATGGCGAAGGCATGGAAGGCCTGTTTTGCCTGCATGAAAGCGAAATTGAAGCG
CTGGCGGAAGATGAAATGTGGAAAGAACTGGTGGAATATATTGGC

Tran6 Protein Sequence

MGIFEHVQDVIGQLPVLKSYSHMLICFPVQDDKREVAIQELERAVRLVMKTFPYLSGKVI
NEGSGAGSSGTFKVTSYKEWESETHVFVRVQDRTTECPAYDELCAARGPSSMLPGH
LLSSRVAFPETYQDIEEDPAPVLDFQANIVRGGLLLDLAAQHNIIDGTGIFQIINLLATALR
GDQFPLFQLHEGNRDRRDLIRLLGPDEALLDHSELKPPVIMKAPPPSDILAPYKWRYYR
FPVDSVNKIRDLANSKPEDFDPSTESLSLNDAITAFCWQRITAIRLRKLKTPTAFSKLSR
AVDFRRIMRLTPAYLGHMVRVCNTRLTFEDIVESSLSRLASLLRKDVQEISNEYALRSY
VTFIANEPDKSDIAYGGSFNAQTDFSCSSIAHVKAPDFGPLGKPGLMRRPTFQPLPCSS
YIAPMLHGEGMEGLFCLHESEIEALAEDEMWKELVEYIG

Codon optimized *tran7* sequence

ATGGCGTTTCAGGATGTGCTGGGCCAGCTGCCGCTGCTGAAAAGCTATACCCATA
TTCTGTTTTGCTTTAGCCTGAGCGATAGCGATCGCGATACCGTGATTGCGGCGCTG
CAGAGCGCGACCCAGCATCTGCTGGCGGCGTTTCCGTTTCTGGGCGGCCAGGTG
GTGCATAAACAGGTGCGCGCGGGCCATAGCGGCTATTTTACCATTGAAGATTGGG
TGGCGCCGAGCCCGACCGATACCCGCGGCGAAGCGGGCAACGCGGAAGAAATTG
AACGCAAAATTCTGCATATTCGCGATCTGAGCGATACCCTGCCGAGCTATGCGGC
GCTGTGCGCGGCGCGCGCGCCGAGCTATATGCTGCCGGGCAGCCTGCTGGCGC

CGCCGCGCCCGGCGTTTCCGCTGACCTATCCGCAGGATGGCCTGGCGCCGGTGC
TGGAAATTCAGGCGAGCCTGATTCAGGGCGGCCTGCTGCTGGATCTGGCGGCGC
AGCATAACATTATTGATGCGACCGGCATTTTTCATATGGCGGGCCTGCTGGCGCG
CCTGATGGATAGCCCGCCGGGCGATATTCAGAAACCGGAAATTCAGCTGGGCAAC
TGCGATCGCCGCAACCTGATTCCGCTGCTGCCGGATCAGGAACCGCTGCCGGATA
ACCTGAGCGTGCTGAAACCGGAACGCTTTCCGCCGCAGCTGGATACCGATACCCT
GGCGCAGTTTAAATGGTATTGGCTGCATTTTAGCCCGGAAGCGATTGTGCATATTT
GCCGCGAAGCGAACAGCCATCCGGAAGATTTTATTGATCCGATTACCAGCGTGAG
CGTGAACGATGCGCTGACCGCGTTTTGCTGGCAGCGCATTGCGCTGATTCGCTTT
CAGACCAGCAGCGGCCCGGATAACGATGAAGAAGAAGAAGAAGAAAAAACCAGCC
AGCTGACCCGCGCGACCGATCTGCGCCGCGCGATGAACCTGAGCCCGGCGTATA
TGGGCCATATGGTGCGCACCGCGAACCTGCGCCTGCCGGTGAGCCTGATTACCC
GCAGCAGCCTGAGCAACCTGGCGAGCCGCCTGCGCGTGTGCGTGCAGGAACATA
CCACCCCGTATGCGGTGCGCGCGTATGCGAGCCTGATTGCGCGCGAAACCGATA
AAAGCCGCATTGCGTATGCGGGCAGCTTTGATCCGCGCACCGATTTTAGCTGCAG
CAGCGTGGCGCATATTAAACTGCCGTTTTTTGGCGGCCTGGGCAAACCGGAATTTA
TGCGCCGCCCGACCTTTGGCCCGCTGCCGGGCGGCATGTATATTGGCCCGGGCT
GGGGCGGCGTGGGCCTGGATGCGGTGGTGTGCCTGCTGGGCACCGAAATGGAA
CTGCTGAAAAAAGATGAATGCTGGAACAGCCTGGTGGAAACCATTGAA

Tran7 Protein Sequence

MAFQDVLGQLPLLKSYTHILFCFSLSDSDRDTVIAALQSATQHLLAAFPFLGGQVVHKQ
VRAGHSGYFTIEDWVAPSPTDTRGEAGNAEEIERKILHIRDLSDTLPSYAALCAARAPS
YMLPGSLLAPPRPAFPLTYPQDGLAPVLEIQASLIQGGLLLDLAAQHNIIDATGIFHMAG
LLARLMDSPPGDIQKPEIQLGNCDRRNLIPLLPDQEPLPDNLSVLKPERFPPQLDTDTL
AQFKWYWLHFSPEAIVHICREANSHPEDFIDPITSVSVNDALTAFCWQRIALIRFQTSSG
PDNDEEEEEEKTSQLTRATDLRRAMNLSPAYMGHMVRTANLRLPVSLITRSSLSNLAS
RLRVCVQEHTTPYAVRAYASLIARETDKSRIAYAGSFDPRTDFSCSSVAHIKLPFFGGL
GKPEFMRRPTFGPLPGGMYIGPGWGGVGLDAVVCLLGTEMELLKKDECWNSLVETIE

Codon optimized *tran8* sequence

ATGCAGGGCACCAACATTTATGAACTGAGCGATCTGGATAAAGGCGGCTTTACCAA
AACCGTGAAAGCGGTGCTGTTTTATGATCTGCATGCGACCGTGGATCCGGAAAGC
GTGGTGAGCAGCCTGCTGGAAGGCGTGAAACATGCGACCCGCCAGCTGCCGTTT
ATGGCGGGCCATCTGGATTTTAACAAAAGCGAACGCCTGTGCATTGTGACCAGCC
CGGATAGCGCGGTGGAAGTGAGCGTGCGCCGCTATACCGGCACCGAATGCGAAC
CGTTTCATGATCTGGCGCAGGATAGCTTTAGCCCGAGCAAACTGGATCTGACCCAT
TTTCTGCCGAAAGATCCGATTAGCAAACATCCGGTGTGCCTGATGCAGATTAACCT
GATTGAAGGCGGCCTGGCGGTGGGCTTTCGCGTGGATCATGCGGCGGGCGATTG
GGTGAGCCTGAGCACCTTTATTAGCCTGGTGTGCCAGAGCAGCAAAGCGCATCAG
GAAGGCCTGGCGATGCCGACCTATACCCCGGATCTGAAACGCGATCCGTATAACG
CGCCGACCCTGGATACCAAAATTAGCCAGCAGGATCGCCTGGCGCAGCTGCCGC
TGTTTCATATTATTGAAAAAAGCAAATTTCAGTTTAAACCGCCGCCGCCGGCGCGC
GCGGGCATTTATCGCATTACCGAACCGACCATTCAGCAGCTGAAAGCGCGCTGCA
TTCCGCATCTGAACCAGGTGGAATATATTACCAGCTATGATTGCATTAGCGCGCTG
CTGTGGCGCGCGCTGACCCGCGCGCGCCTGTATATTCATCCGGATCAGGCGAAC

GCGCCGAGCCGCTTTGTGCATCCGATTGATGTGCGCAGCCGCGATCCGGAACATA
AAACCAGCCCGCAGTATTTTGGCAACGCGGTGATTGGCACCCTGGCGGGCCCGG
TGCCGGCGACCACCCTGATTAGCGAAGGCGATCGCGGCCTGGCGGCGATTGCGA
CCAAAATTCGCCAGAGCATTCATGCGGTGGATATTAGCAAAATTGGCCATCTGACC
GCGCTGCAGACCAGCCTGGCGGATACCGAAATGCTGATTCCGAACGCGGATTTTG
CGGATATGGATCTGTTTATGAACACCTGGTATACCGGCAGCGCGGCGAAATATGAT
CTGGGCGCGGCGACCCGCCCGGTGGCGTTTCGCGTGCAGGCGGGCGTGCCGGG
CGCGTGCGCGATTATTCTGCCGGATCTGAGCGAAAGCGATATGCGCGTGTTTGAA
GTGTATGTGCAGGCGCCGGAAAAAGAACATGATGCGCTGGTGCGCGATGCGGAA
TTTATGAAATATTTTGAACAGGTGGCG

Tran8 Protein Sequence
MQGTNIYELSDLDKGGFTKTVKAVLFYDLHATVDPESVVSSLLEGVKHATRQLPFMAG
HLDFNKSERLCIVTSPDSAVEVSVRRYTGTECEPFHDLAQDSFSPSKLDLTHFLPKDPI
SKHPVCLMQINLIEGGLAVGFRVDHAAGDWVSLSTFISLVCQSSKAHQEGLAMPTYTP
DLKRDPYNAPTLDTKISQQDRLAQLPLFHIIEKSKFQFKPPPPARAGIYRITEPTIQQLKA
RCIPHLNQVEYITSYDCISALLWRALTRARLYIHPDQANAPSRFVHPIDVRSRDPEHKTS
PQYFGNAVIGTLAGPVPATTLISEGDRGLAAIATKIRQSIHAVDISKIGHLTALQTSLADTE
MLIPNADFADMDLFMNTWYTGSAAKYDLGAATRPVAFRVQAGVPGACAIILPDLSESD
MRVFEVYVQAPEKEHDALVRDAEFMKYFEQVA

Codon optimized *tran9* sequence
ATGAGCATTGCGATTGAAAACCATCTGACCAAAGGCGAAACCATTGAAACCGATAG
CGAACTGCCGCGCGATATTGTGAGCCAGTTTCCGATTCTGAACGCGTATACCCAG
CTGCTGTTTGGCTTTAAACTGCCGGCGGATGTGGATCGCGATGCGATTGTGGCGA
GCCTGCAGGATGGCTTTGATAAACTGCGCGCGGAAATTCCGTGGCTGGGCTGGCA
GGTGGCGCGCGAAAGCGGCCCGAACGGCATTCTGAAAGCGATGCCGTGGCCGG
CGGATGTGCCGAAAGAACGCATTCGCGTGAAAAACTGCGATGATCTGGTGGCGCC
GATTAGCAAACTGGTGAGCGCGGGCGTGCCGATTCATATGCTGGATGGCAGCGTG
CTGACCCCGTGGCCGGCGCTGCCGCAGCCGCGCGGCCTGGATAGCCCGGATCC
GGTGGTGGCGATGCAGGCGAACTTTGTGCAGGGCGGCCTGATTCTGAACCTGAG
CACCCATCATACCATTATTGATGGCACCGGCATTTATCAGTTTGTGAACCTGCTGG
CGCTGGTGATGAGCGGCAAAAAAATTCCGGCGGCGGATCTGGAACAGGCGAACC
GCGATCGCAGCCGCGTGATTCCGCTGATTCCGCGCACCGAACCGGTGAAAAGCTA
TATTCATCTGCGCCGCCCGCCGGGCTATACCTGGGCGCCGCCGAGCAGCCCGCC
GATGTGGTGCTATTTTAAAATTCCGGTGAACGCGCTGGCGCGCCTGATTAAAAGC
GTGAAAGATGATCCGCTGAGCAACAAACCGGGCAGCATGATGGTGAGCGATAACG
ATATTTTTAGCGCGTTTGCGTGGCAGCGCCTGTGCGCGGTGCGCGTGGCGAACG
GCCAGCCGCCGGAACGCAGCAGCAAACTGGGCCGCGCGATTGATGGCCGCATGG
CGCTGGGCGTGCCGCTGACCTATATGGGCCATATGGTGTGCCATGCGCTGGTGC
GCCTGCCGCTGGGCCAGGTGGCGAAACTGCCGCTGCCGCAGCTGGCGCAGGTG
CTGCGCCGCGAACTGAACCAGGCGAACACCCCGTGGAGCATTCGCAGCTATGCG
ACCTTTATGGCGCGCGAACCGGATACCAGCAGCCTGCTGTATGGCGGCACCCATG
ATGCGCGCGCGGATCTGATGGTGACCGCGGTGGGCCAGGCGACCCCGCTGCCG
GCGAGCTGGGGCCCGCTGCTGGGCCGCAGCTGCTTTTTTCGCCGCCCGACCGCG
GCGCCGATTCCGGGCTGCTTTATTATTAACGAAACCGAAGGCGCGTTTATTCCGCT

GACCCTGTGCATGCCGGAAGAAGATATTATTGGCCTGAAAAAAGATCCGGTGTGG
AAACAGTATATTCGCTATGTGGGC

Tran9 Protein Sequence

MSIAIENHLTKGETIETDSELPRDIVSQFPILNAYTQLLFGFKLPADVDRDAIVASLQDG
FDKLRAEIPWLGWQVARESGPNGILKAMPWPADVPKERIRVKNCDDLVAPISKLVSAG
VPIHMLDGSVLTPWPALPQPRGLDSPDPVVAMQANFVQGGLILNLSTHHTIIDGTGIYQ
FVNLLALVMSGKKIPAADLEQANRDRSRVIPLIPRTEPVKSYIHLRRPPGYTWAPPSSP
PMWCYFKIPVNALARLIKSVKDDPLSNKPGSMMVSDNDIFSAFAWQRLCAVRVANGQ
PPERSSKLGRAIDGRMALGVPLTYMGHMVCHALVRLPLGQVAKLPLPQLAQVLRREL
NQANTPWSIRSYATFMAREPDTSSLLYGGTHDARADLMVTAVGQATPLPASWGPLLG
RSCFFRRPTAAPIPGCFIINETEGAFIPLTLCMPEEDIIGLKKDPVWKQYIRYVG

Codon optimized *tran10* sequence

ATGAGCGGCCTGCCGCGCGGCGTGGATGTGCTGGGCCAGTTTCCGCAGCTGAAA
ACCTATAACCATGGCGTGGCGCTGTTTCCGTGGCGCGATGATCTGACCCAGGAAA
TTGTGATTAGCGCGCTGCGCGATGCGACCGGCAAAATTCTGAACGCGATTCCGTG
GCTGACCGGCCAGGTGGTGCATACCGTGGGCGCGAACGGCACCAGCGGCACCAT
TCGCGTGGAACCGTGGCCGCCGGGCGTGCCGCCGCATGGCCTGCTGAAAATTAA
AGATGCGACCGATCTGCTGCCGCCGTATGCGGATATGTGCGCGGCGAACGCGCC
GTTTTATATGCTGGATGGCAAAGTGCTGTGCCCGTATCCGGGCTTTCCGGAACGCT
ATGATACCGATGTGATTGGCCCGGCGCCGCCGGCGATGATTCAGGTGACCTTTGT
GAAAGGCGGCTTTCTGCTGACCTATACCAACCAGCATAACGTGGTGGATGGCACC
GGCCTGTTTACCATTATTACCCTGCTGAGCTTTGCGCTGCGCGGCGAACCGTTTCC
GGAAGATATTATTGCGGCGGCGAACCAGGATCCGGCGGGCGTGATTCGCCTGAT
GGATCCGTTTGAACCGCTGCTGGATCATAGCTGGCTGCTGAAACCGGATCCGGAA
CCGGAATTTGTGCCGGCGCGCCCGGGCGGCCGCTGGCAGAACTTTCTGTTTCGC
AAAGAAGCGGTGGCGGATGTGAAAAAAGCGGCGGAAGATCCGATTGGCTATGATG
AAGAAGTGCCGTTTATTAGCCCGAACGATGCGATTTGCGCGCTGTTTTGGAAAAGC
CTGGCGAGCGTGCGCATTAAAACCGGCCTGGCGCCGACCACCGTGAGCAAATTTA
ACCGCGCGATTGATACCCGCCCGGTGCTGGGCGTGAGCCCGGGCTATCTGGGCC
AGATGGTGTATATTAGCCCGAGCCGCATGACCTTTCAGGAACTGGTGCATGCGCC
GCTGAGCACCGTGACCACCCGCCTGCGCAAAGATCTGGATGATGTGAACAACGAA
TTTAGCGCGCGCAGCTATGCGACCTTTGTGGCGAGCGTGGATGATAAAACCCAGC
TGCTGTATAGCGGCGGCATGAACCTGCAGACCGATGTGGGCCATAGCAGCATGAT
GAACGCGAAAGTGAACGGCTTTGATTATGGCATTCTGGGCAAAAGCGATTTTGGC
CGCCGCCCGCATCTGGGCCTGACCGTGAACACCTTTTATATTTATCCGCAGGAAAA
AAGCGGCGATTATCTGATTCTGGTGTGCCTGCCGGATATGGAACTGCAGGGCCTG
ATGGAAGATCCGGATTGGGCGCGCTGCACCTATGCGATTCAGGAAGAATAT

Tran10 Protein Sequence

MSGLPRGVDVLGQFPQLKTYNHGVALFPWRDDLTQEIVISALRDATGKILNAIPWLTGQ
VVHTVGANGTSGTIRVEPWPPGVPPHGLLKIKDATDLLPPYADMCAANAPFYMLDGKV
LCPYPGFPERYDTDVIGPAPPAMIQVTFVKGGFLLTYTNQHNVVDGTGLFTIITLLSFAL
RGEPFPEDIIAAANQDPAGVIRLMDPFEPLLDHSWLLKPDPEPEFVPARPGGRWQNFL
FRKEAVADVKKAAEDPIGYDEEVPFISPNDAICALFWKSLASVRIKTGLAPTTVSKFNRA

IDTRPVLGVSPGYLGQMVYISPSRMTFQELVHAPLSTVTTRLRKDLDDVNNEFSARSY
ATFVASVDDKTQLLYSGGMNLQTDVGHSSMMNAKVNGFDYGILGKSDFGRRPHLGLT
VNTFYIYPQEKSGDYLILVCLPDMELQGLMEDPDWARCTYAIQEEY

Codon optimized *tran11* sequence

ATGACCATGGATAACACCCAGGTGCAGGCGGGCGATGATGATCTGTATCAGGATG
TGATGAGCCAGTTTCCGTTTCTGAACGGCTATACCCATTTTATGCTGGGCTTTCAG
CTGGATAGCGATGCGAGCTGCAACGCGATTGTGGATAGCCTGAAAACCGGCGTGC
AGCGCGTGATTGATAAAGTGCCGTGGCTGGGCGGCCAGGTGGCGTGCGAAAGCA
GCCCGGAAGGCACCGCGACCTTTCGCCCGACCGCGTGGCCGGATGGCCTGCCG
GTGAACGAAATTGTGCGCCTGAACATTTGCGATGCGGAACTGCCGCCGATGGCGG
AACTGCTGCGCACCGGCACCCCGGTGAGCGTGCTGCCGGCGAGCGTGCTGACCC
CGTGGCCGGGCCTGCCGCAGCCGCATGGCATTAGCGGCCCGGTGCCGGTGCTG
GCGATGCAGGCGAACTTTATTAGCGGCGGCCTGATTCTGACCATTAGCTGCCATC
ATAACATTATGGATGCGACCGCGATTTTTCAGTTTATTCGCCTGCTGACCGTGGTG
ATGAACAACGGCGAAATTATGACCCCGGAACTGGATCAGGCGAACCGCGATCGCC
GCGGCGTGGTGCCGCTGATTCCGCGCGGCGAACCGGTGAAAAACTTTAGCCATCT
GCGCCGCGGCCCGGATTATGTGCAGCGCACCCCGCCGAGCACCAGCCGCTGGT
GCTATTTTCGCCTGCCGGTGGCGGCGATGCCGCGCATGCGCAAACTGGCGAGCC
GCCCGCAGCGCACCATTAGCGATCAGGGCAGCCCGGCGCCGCTGCTGAGCGATA
ACGATATTCTGAGCGCGTTTTGCTGGAACCGCATTAGCGCGGGCCGCCTGGGCCG
CGGCGCGATGACCGGCGGCCAGAAAGTGGGCCCGGCGAGCACCACCAAATTTAA
CCGCGCGATTGATGGCCGCGTGGCGCTGGGCGTGCCGCCGAGCTATATGGGCCA
TATGGTGTGCCATGCGAGCACCCGCCTGACCCTGCAGCAGGTGGTGACCCAGCC
GCTGAGCACCACCGCGGCGGCGCTGCGCCGCGATCTGAAGCGGCGAACACCCC
GTGGAGCATTCGCAGCTATGCGACCTTTCTGGCGCGCGAACCGGATCGCAGCAAA
CTGCTGTATGGCGGCAGCTTTGATCCGAACACCGATATTGGCGCGAGCAGCCTGG
TGACCGATCTGACCGGCGCGAGCCAGGATGATACCGCGAGCGGCTTTGCGTTTAT
GAGCCTGGGCGATCTGCTGGGCAAACCGCGCTTTATTCGCCGCCCGGATCTGCC
GCTGATTCCGGGCGTGCTGTATTTTCAGCCGGTGGAAAACGGCGCGATTCCGATT
CTGCTGTGCCTGACCGGCGAAGATCTGGATGGCCTGAAAAAAGATCGCCAGTGGA
CCCAGTATATGGAATATGTGGGC

Tran11 Protein Sequence

MTMDNTQVQAGDDDLYQDVMSQFPFLNGYTHFMLGFQLDSDASCNAIVDSLKTGVQ
RVIDKVPWLGGQVACESSPEGTATFRPTAWPDGLPVNEIVRLNICDAELPPMAELLRT
GTPVSVLPASVLTPWPGLPQPHGISGPVPVLAMQANFISGGLILTISCHHNIMDATAIFQ
FIRLLTVVMNNGEIMTPELDQANRDRRGVVPLIPRGEPVKNFSHLRRGPDYVQRTPPS
TSRWCYFRLPVAAMPRMRKLASRPQRTISDQGSPAPLLSDNDILSAFCWNRISAGRLG
RGAMTGGQKVGPASTTKFNRAIDGRVALGVPPSYMGHMVCHASTRLTLQQVVTQPL
STTAAALRRDLNAANTPWSIRSYATFLAREPDRSKLLYGGSFDPNTDIGASSLVTDLTG
ASQDDTASGFAFMSLGDLLGKPRFIRRPDLPLIPGVLYFQPVENGAIPILLCLTGEDLDG
LKKDRQWTQYMEYVG

Codon optimized *tran12* sequence

ATGACCATTGATAACACCCAGGCGCAGGTGGGCGATGATGATCTGTATCAGGATG
TGATGAGCCAGTTTCCGTTTCTGAACGGCTATACCCATTTTATGCTGGGCTTTCAG
CTGGATAGCGATGAAAGCCATAACGCGATTGTGGATGCGCTGAAAACCGGCGTGC
AGCGCGTGATTGATAAAGTGCCGTGGCTGGGCGGCCAGGTGGCGTGCGAAAGCA
GCCCGGAAGGCACCGCGACCTTTCGCCCGACCGCGTGGCCGGAAGGCCTGCCG
GTGAACGAAATTGTGCGCCTGAACATTTGCGATGCGAAACTGCCGCCGATGGCGG
AACTGCTGCGCACCGGCACCCCGGTGAGCGTGCTGCCGGCGAGCGTGCTGACCC
CGTGGCCGGGCCTGCCGCAGCCGCATGGCATTAGCGGCCCGGTGCCGATTATTG
CGATGCAGGCGAACTTTATTAGCGGCGGCCTGATTCTGACCATTAGCTGCCATCAT
AACATTATGGATGCGACCGCGATTTTTCAGTTTATTCGCCTGCTGACCGTGGTGAT
GAAAAACGGCGAAATTATGCCGCCGGAACTGGATCAGGCGAACCGCGATCGCCG
CGGCGTGATTCCGCTGATTCCGCGCGGCGAACCGGTGAAAAACTTTAGCCATCTG
CGCCGCGCGCCGGATTATGTGCAGCGCACCCCGAGCAGCACCAGCCGCTGGTGC
TATTTTCGCTTTCCGGTGGCGGCGATGCCGCGCATGCGCAAACTGGCGAGCCGCC
CGCAGACCACCATGAGCGAACAGGGCAGCCCGGCGCCGCTGCTGAGCGATAACG
ATATTCTGAGCGCGTTTTGCTGGAACCGCATTAGCGCGGTGCGCCTGGGCCGCGG
CGCGATGACCGGCGGCCAGAAAGTGGGCCCGACCAGCACCACCAAATTTAACCG
CGCGATTGATGGCCGCGTGGCGCTGGGCGTGCCGCCGAGCTATATGGGCCATAT
GGTGTGCCATGCGAGCACCCGCCTGACCCTGCAGCAGGTGGTGACCCAGCCGCT
GAGCACCACCGCGGCGGCGCTGCGCCGCGATCTGAACGCGGCGAACACCCCGT
GGAGCATTCGCAGCTATGCGACCTTTCTGGCGCGCGAACCGGATCGCAGCAAACT
GCTGTATGGCGGCAGCTTTGATCCGAACACCGATATTGGCGCGAGCAGCCTGGTG
ACCGATCTGACCGGCGCGAGCAAAGATGATACCGCGAGCGGCTTTGCGTTTATGA
GCCTGGGCGATCTGCTGGGCAAACCGCGCTTTATTCGCCGCCCGGATCTGCCGCT
GATTCCGGGCGTGCTGTATTTTCAGCCGGTGGAAAACGGCGCGATTCCGATTCTG
CTGTGCCTGAGCGCGGAAGATCTGGATGGCCTGATGAAAGATCGCCAGTGGACC
CAGTATATGAAATATGTGGGC

Tran12 Protein Sequence

MTIDNTQAQVGDDDLYQDVMSQFPFLNGYTHFMLGFQLDSDESHNAIVDALKTGVQR
VIDKVPWLGGQVACESSPEGTATFRPTAWPEGLPVNEIVRLNICDAKLPPMAELLRTG
TPVSVLPASVLTPWPGLPQPHGISGPVPIIAMQANFISGGLILTISCHHNIMDATAIFQFIR
LLTVVMKNGEIMPPELDQANRDRRGVIPLIPRGEPVKNFSHLRRAPDYVQRTPSSTSR
WCYFRFPVAAMPRMRKLASRPQTTMSEQGSPAPLLSDNDILSAFCWNRISAVRLGRG
AMTGGQKVGPTSTTKFNRAIDGRVALGVPPSYMGHMVCHASTRLTLQQVVTQPLSTT
AAALRRDLNAANTPWSIRSYATFLAREPDRSKLLYGGSFDPNTDIGASSLVTDLTGASK
DDTASGFAFMSLGDLLGKPRFIRRPDLPLIPGVLYFQPVENGAIPILLCLSAEDLDGLMK
DRQWTQYMKYVG

Overexpression of AT enzymes and preparation of clarified cell lysate: Plasmid containing the sequences of interest were used to transform chemically competent *E. coli* BL21(DE3) cells using standard heat-shock protocols. Overexpression of the proteins

was achieved in 1 mL of Luria Broth (LB) in 1.5 mL plastic tubes. 1 mL portions of media were inoculated with a single colony from an agar plate, along with 50 μg/mL kanamycin (Gold Biotechnology). Cultures were grown at 37 °C and 200 rpm until the optical density at 600 nm reached 0.8 (approx. 3.5 h). The cultures were then cooled to 16 °C for 30 min and expression was induced with 0.1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG, Gold Biotechnology). Expression continued at 16 °C overnight (approx. 18 h) at 200 rpm. Cultures were then pelleted by centrifugation at 17,000 xg for 10 min and the supernatant was removed. Cell pellets were resuspended in 50 mM KPi buffer pH 8.0 at a

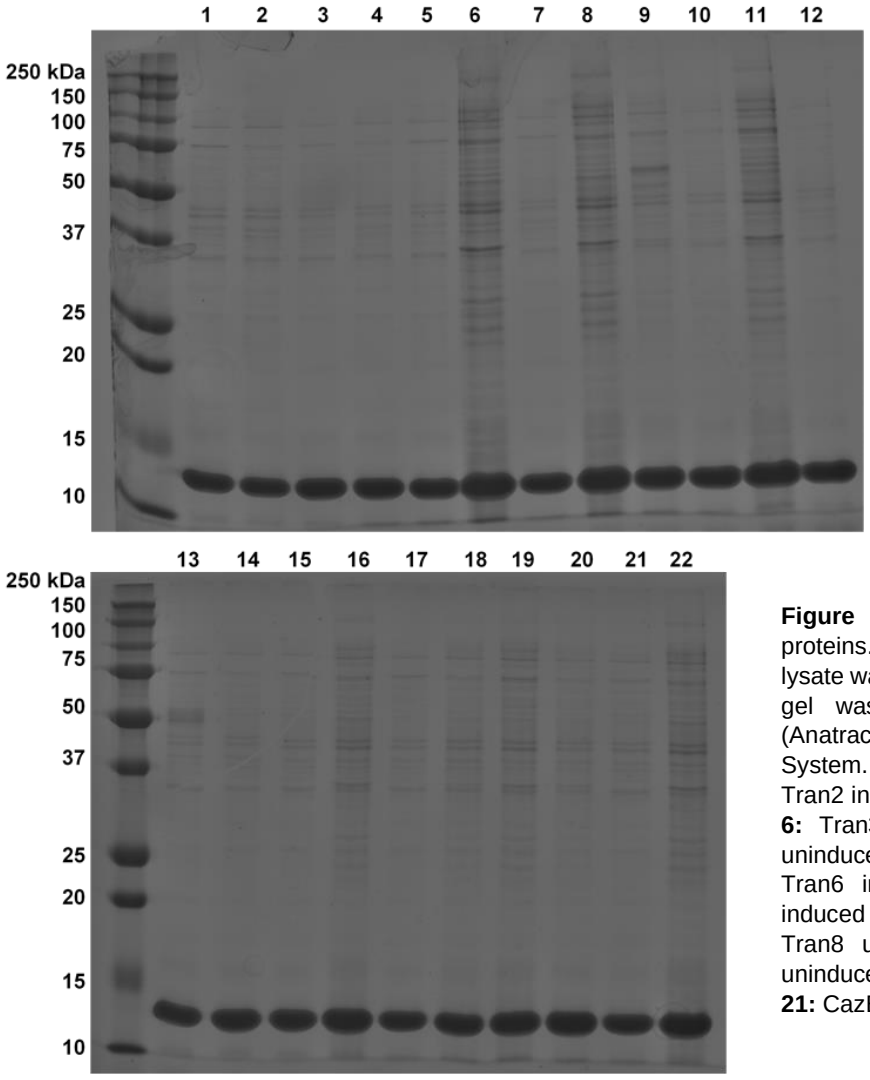


Figure 4.6. Induced and uninduced transferase proteins. Approximately 7.5 µL each protein clarified lysate was loaded onto a 12.5% SDS-PAGE gel. The gel was stained with Quick Coomassie stain (Anatrace) and visualized with the Azure Gel Imaging System. **1:** Tran1 induced **2:** Tran1 uninduced **3:** Tran2 induced **4:** Tran2 uninduced **5:** Tran3 induced **6:** Tran3 uninduced **7:** Tran4 induced **8:** Tran4 uninduced **9:** Tran5 induced **10:** Tran5 uninduced **11:** Tran6 induced **12:** Tran6 uninduced **13:** Tran7 induced **14:** Tran7 uninduced **15:** Tran8 induced **16:** Tran8 uninduced **17:** Tran9 induced **18:** Tran9 uninduced **19:** Tran10 induced **20:** Tran10 uninduced **21:** CazE induced **22:** CazE uninduced

concentration of 50 mg of cell pellet/mL. The cell suspensions were then lysed by sonication on ice, 10 sec on 20 sec off x3. Insoluble material was pelleted by centrifugation at 17,000 xg for 10 min and the clarified supernatant was used directly for biocatalytic reactions.

Large-scale PigD overexpression and purification: Plasmid containing the *pigD sequence* was used to transform chemically competent *E. coli* BL21(DE3) cells using standard heat-shock protocols. Overexpression of the protein was achieved in 1 L of Terrific Broth (TB) in 2.8 L flasks. 1 L portions of media were inoculated with 10 mL overnight culture prepared from a single colony in Luria Broth (LB) and 50 μg/mL kanamycin (Gold Biotechnology). Cultures were grown at 37 °C and 200 rpm until the optical density at 600 nm reached 0.8 (approx. 3.5 h). The cultures were then cooled to 16 °C for 30 min and expression was induced with 0.1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG, Gold Biotechnology). Expression continued at 16 °C overnight (approx. 18 h) at 200 rpm. The typical yield for 1 L culture was ~17 g cell pellet.

PigD purification procedure: Approximately 200 g of cell pellet was resuspended in 600 mL of lysis buffer containing 50 mM Tris HCl pH 7.5. 1 mg/mL lysozyme was added to the resuspended cells that were then incubated on a rocker at 4 °C for 30 min. Cells were lysed by first sonicating the mixture on ice, 10 sec on 20 sec off for a total of 3 min on. The total cell lysate was then passed through an Avestin pressure homogenizer at 15000 psi. The total lysate was centrifuged at 40,000 x g for 30 min, then the supernatant was mixed with roughly 5 mL of concentrated Ni-NTA resin and incubated on a rocker at 4 °C for 1h. The impregnated resin was removed from the mixture by a gravity flow column

and rinsed with 30 mL of wash buffer containing 50 mM Tris HCl pH 7.5 and 5 mM imidazole. The protein was eluted from the resin with 10 mL of elution buffer containing 50 mM Tris HCl pH 7.5 and 250 mM imidazole. The eluted protein was passed through a desalting PD10 column and eluted in a storage buffer containing 25 mM Tris HCl pH 8.4 and 10% glycerol. The eluted protein was then concentrated to 1.7 mM (determined by Pierce assay) using a 30,000 NMWL Amicon Ultra-15 centrifugal filter. Average yields: 14 mg from 1 L PigD. The molecular weight of PigD including the 6xHis-tag was estimated by the ProtParam tool on the Expasy server to be 50.2 kDa. This molecular weight is consistent with the mass of the protein band observed by SDS-PAGE analysis (Figure 4.6). The purified protein was aliquoted into 0.6 mL tubes and frozen in liquid nitrogen before long-term storage at -80 °C.

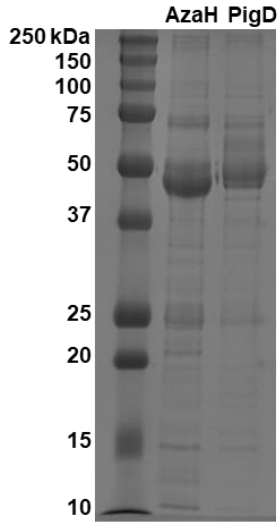


Figure 4.7. Purified AzaH and PigD. Approximately 2 µL of 20 µM each protein was loaded onto a 12.5% SDS-PAGE gel. The gel was stained with Quick Coomassie stain (Anatrace) and visualized with the Azure Gel Imaging System. The relative apparent masses are consistent with the predicted estimates.

III. Biocatalytic reactions

Stock solutions: Stock solutions of each substrate (50 mM) were prepared by dissolving the substrate in DMSO (analytical grade) and stored at -20 °C. Stock solutions of $NADP^+$ (100 mM) and glucose-6-phosphate (G6P, 500mM) were stored at -20 °C. Aliquots of each enzyme and glucose-6-phosphate dehydrogenase (G6PDH, 100 U/mL) were stored at -80 °C.

Analytical-scale reactions with clarified cell lysate: Each reaction contained 2.5 mM of bicycle **4.19** substrate (5 uL of a 50 mM stock solution in DMSO), 2.5 mM thioester **4.20** (5 uL of a 50 mM stock solution in DMSO), in enzyme clarified lysate (90 uL, prepared as described above). The reactions were quenched with 200 µL of MeOH and the precipitated biomolecules were pelleted by centrifugation (16,000 x g, 12 min). The supernatant was then analyzed by UPLC-DAD and LC-MS.

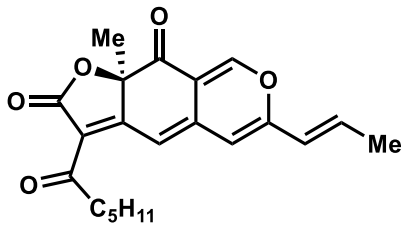
Analytical-scale reactions with purified enzyme: Each reaction contained 2.5 mM of the orcinaldehyde substrate (5 uL of a 50 mM stock solution in DMSO), 5 mM G6P (1 µL, 500 mM), 1 mM NADP+ (1 µL, 100 mM), 1 U/mL G6P-DH (1 µL, 100 U/mL), 5 µM AzaH (3.3 µL of a 150 µM stock solution), and 85 µL of a 50 mM potassium phosphate buffer, pH 8.0. The dearomatization reaction was carried out at 30 °C for 1 h. Following this, the pH of the reactions were adjusted to ~8.4 using 1 µL 200 mM NaOH. 2.5 mM thioester **4.20** (2.5 µL of a 100 mM stock solution in DMSO) and 20 µM PigD (1.2 µL of a 1.7 mM stock solution) were then added, and the second reaction was carried out at 30 °C for 1

h. The reactions were quenched with 200 µL of MeOH and the precipitated biomolecules were pelleted by centrifugation (16,000 x g, 12 min). The supernatant was then analyzed by UPLC-DAD and LC-MS.

General procedure for preparative scale *in vitro* reactions:

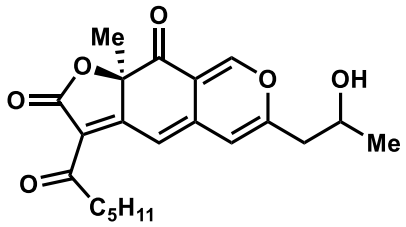
2.5 mM orcinaldehyde substrate (171 uL of a 50 mM stock solution in DMSO), 5 mM G6P (34 µL, 500 mM), 1 mM NADP+ (34 µL, 100 mM), 1 U/mL G6P-DH (34 µL, 100 U/mL), 5 µM AzaH (113 µL of a 150 µM stock solution) was added to 2.8 mL of 50 mM potassium phosphate buffer, pH 8.0. The reaction was placed in an incubator at 30 °C with 85 rpm shaking. After 1 h, the pH of the mixture was adjusted to ~8.4 with 6.8 µL 1 M NaOH. 2.5 mM thioester **4.20** (85 µL of a 100 mM stock solution in DMSO) and 20 µM PigD (40 µL of a 1.7 mM stock solution) were then added, and the reaction was returned to the 30 °C for 1 h without shaking. The reaction was then diluted with a saturated sodium sulfate solution (~10x reaction volume) and extracted 3 times with ~10 mL 9:1 EtOAc:n-BuOH. The combined organic layers were washed with ~30 mL of brine, then dried with sodium sulfate. The sodium sulfate and precipitated biomolecules were removed through gravity filtration and the flow-through was passed through a 0.22 µm filter before the solvent was removed using a rotary evaporator. The resulting crude oil was dissolved in ~500 µL of HPLC grade MeOH and subjected to purification by preparative HPLC using a Phenomenex Kinetex 5 µm C18, 150 x 21.2 mm column under the following conditions: mobile phase A = deionized water and B = acetonitrile; method = 5% to 20% B over 4 min, 20% to 50% B over 9 min, 50% to 100% B over 2 min, 100% B for 3 min; flow rate,

10 mL/min. Fractions containing the purified compound were pooled and the solvent was removed using a rotary evaporator.



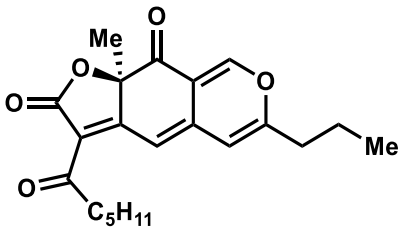
rubropunctatin (4.10)

Following preparation by the general method described above, the title compound was isolated as an orange, amorphous solid (11% yield by NMR over two steps). **^{1}H NMR** (600 MHz, CDCl3) δ 7.86 (s, 1H), 6.89 (s, 1H), 6.56-6.62 (m, 1H), 6.14 (s, 1H), 6.03-6.06 (d, *J* = 7 Hz, 1H), 2.89-2.99 (m, 2H), 1.95-1.96 (d, *J* = 7.0 Hz, 3H), 1.71 (s, 3H), 1.59-1.64 (m, 4H), 1.31-1.33 (m, 2H), 0.88-0.89 (t, *J* = 5.0 Hz, 3H). **^{13}C NMR** (151 MHz, CDCl3) δ 197.6, 191.0, 171.8, 169.4, 156.54, 152.9, 141.7, 136.5, 122.5, 116.5, 113.4, 109.7, 104.3, 85.9, 41.8, 31.6, 28.5, 23.6, 18.9, 14.1. **HR-ESI-MS**: *m/z* calculated for $C_{21}H_{23}O_5$ $[M+H]^+$: 355.1540, found: 355.1546. All spectral data is consistent with available published values for rubropunctatin (**4.10**) and monascorubrin (**4.1**).[48]



(9a*R*)-3-hexanoyl-6-(2-hydroxypropyl)-9a-methyl-2H-furo[3,2-g]isochromene-2,9(9aH)-dione (4.25A)

Following preparation by the general method described above, the resulting red oil was warmed to 40 °C under vacuum overnight (~14 h) to give the pure title compound as a brown oil (20% yield by NMR over two steps). **^{1}H NMR** (600 MHz, CDCl3) δ 7.85 (s, 1H), 6.84 (s, 1H), 6.26 (s, 1H), 2.89-2.98 (m, 2H), 2.55-2.57 (m, 2H), 1.71 (s, 3H), 1.60-1.62 (m, 4H), 1.32-1.33 (m, 2H), 0.87-0.90 (t, *J* = 6.7 Hz, 3H). **^{13}C NMR** (151 MHz, CDCl3) δ 197.4, 190.9, 171.9, 169.1, 160.3, 153.1, 140.9, 116.7, 113.7, 111.6, 111.6, 103.5, 85.90 65.5, 65.4, 42.7, 42.6, 41.6, 31.38, 3.1, 23.7, 23.4, 22.5, 14.0. **HR-ESI-MS**: *m/z* calculated for $C_{21}H_{25}O_6$ $[M+H]^+$: 373.1646, found: 373.1652.

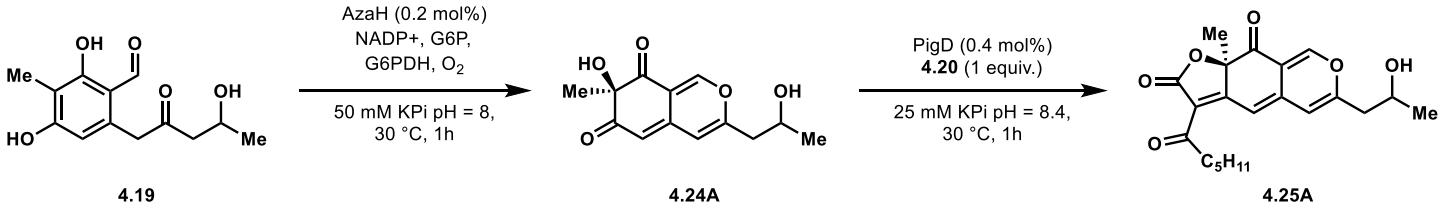


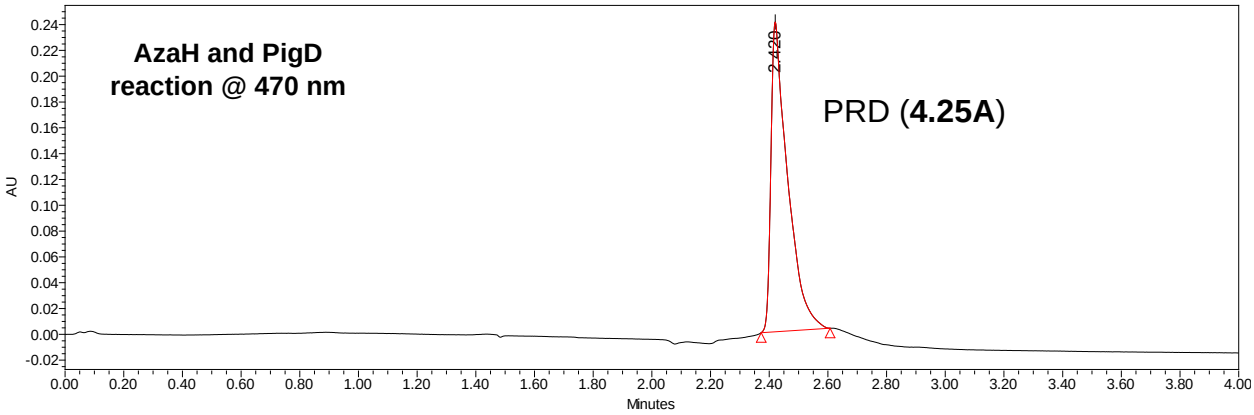
(*R*)-3-hexanoyl-9a-methyl-6-propyl-2H-furo[3,2-g]isochromene-2,9(9aH)-dione (4.25B)

Following preparation by the general method described above, the title compound was isolated as an orange, amorphous solid (20% yield by NMR over two steps). **^{1}H NMR** (600 MHz, CDCl3) δ 7.85 (s, 1H), 6.82 (s, 1H), 6.15 (s, 1H), 2.89-2.98 (m, 2H), 2.40-2.43 (t, *J* = 7.5 Hz, 2H), 1.71 (s, 3H), 1.65-1.69 (m, 4H), 1.31-1.33 (m, 4H), 0.99-1.01 (t, *J* = 7.4, 3H), 0.88-0.90 (t, *J* = 7.0 Hz, 3H). **^{13}C NMR** (151 MHz, CDCl3) δ 197.6, 191.2, 172.3, 169.4, 163.6, 153.5, 141.7, 116.8, 113.4, 109.9, 103.2, 86.0, 77.4, 77.2, 77.0, 41.8, 35.3, 31.53, 28.4, 23.6, 22.7, 20.10, 14.1, 13.6. **HR-ESI-MS**: *m/z* calculated for $C_{21}H_{25}O_6$ $[M+H]^+$: 357.1697, found: 357.1703.

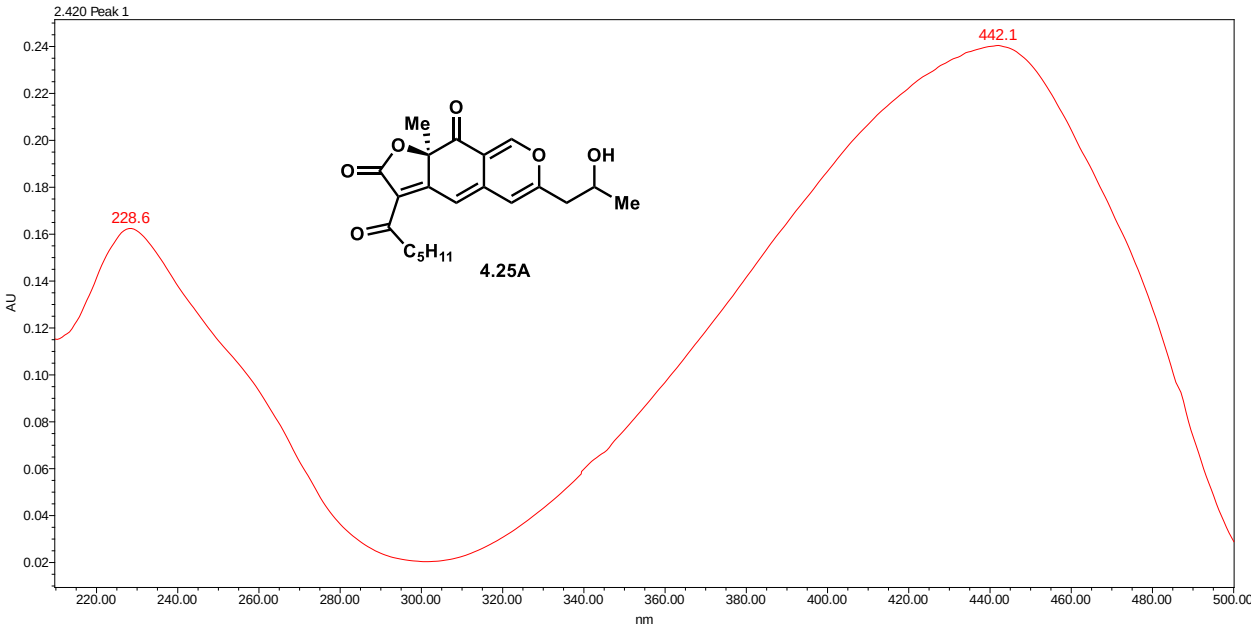
IV. UPLC-DAD and LC-MS Traces

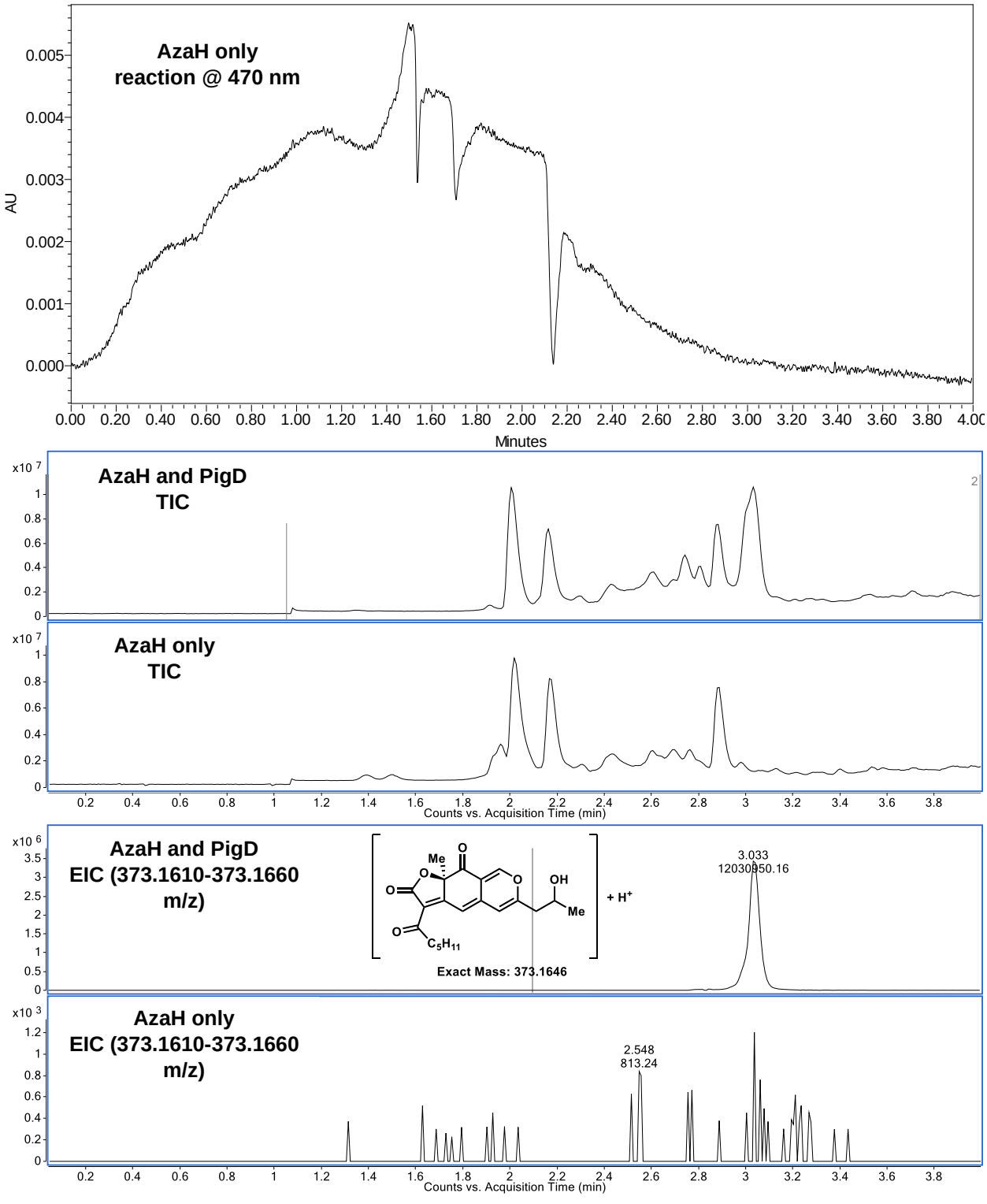
Figure 4.8. One-pot enzymatic transformation of **4.19** by AzaH and PigD. PDA and LC-MS traces of enzymatic reaction and control reaction.





	Retention Time (min)	Area	% Area	Height
1	2.420	949812	100.00	240098




AzaH only
reaction @ 470 nm
AU
Minutes
AzaH and PigD
TIC
AzaH only
TIC
Counts vs. Acquisition Time (min)
AzaH and PigD
EIC (373.1610-373.1660
m/z)
Me
O
OH
Me
+ H+
C5H11
Exact Mass: 373.1646
3.033
12030950.16
AzaH only
EIC (373.1610-373.1660
m/z)
2.548
813.24
Counts vs. Acquisition Time (min)

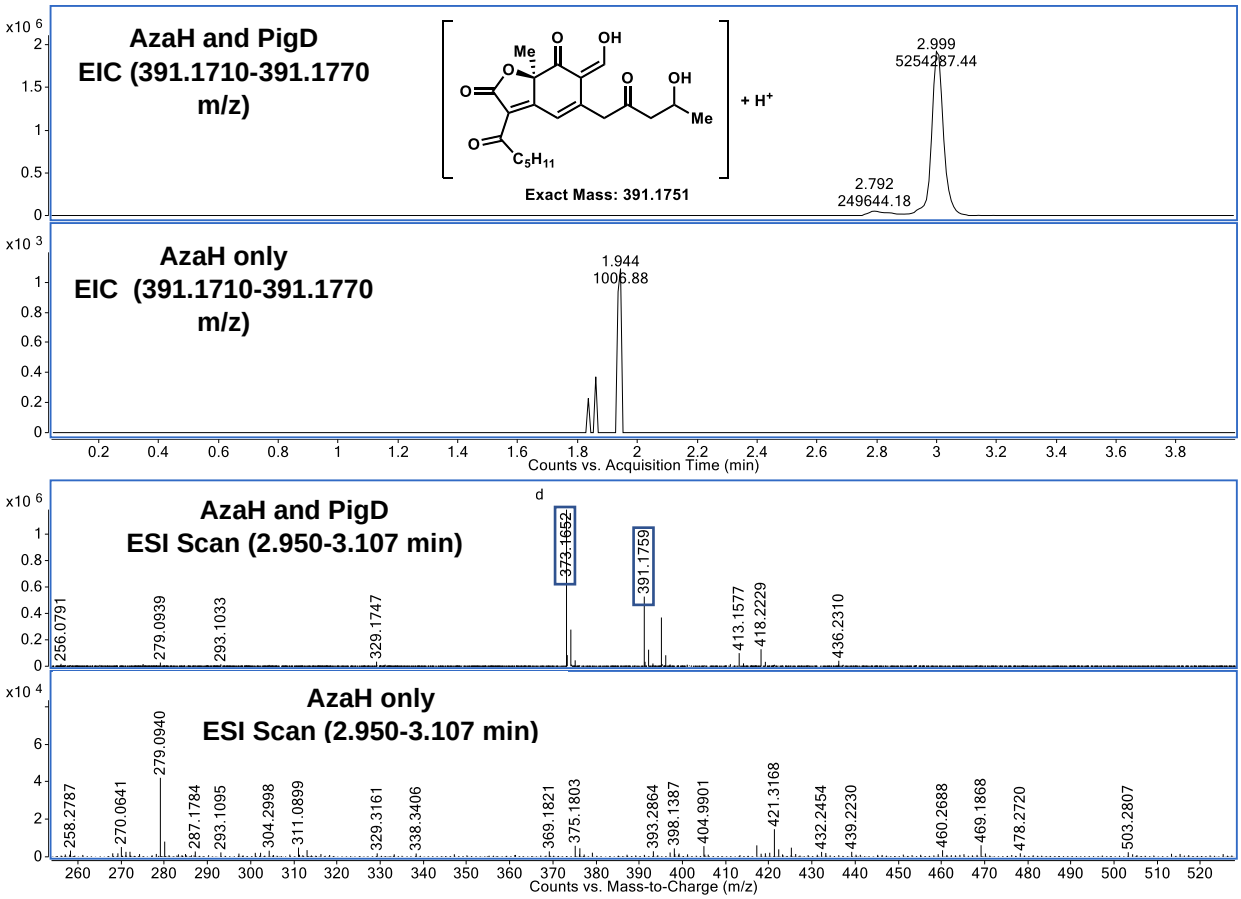
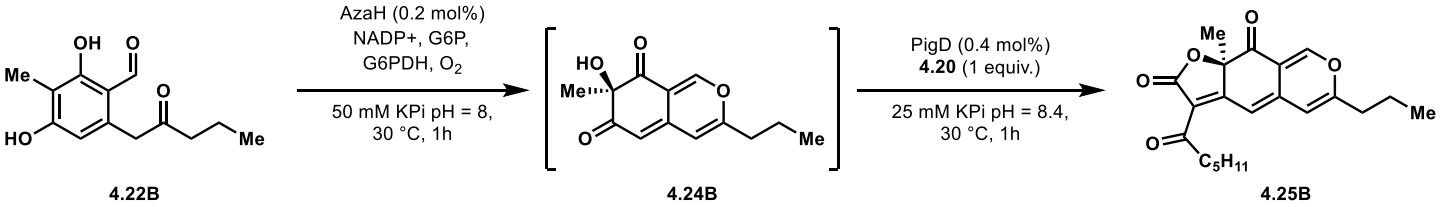
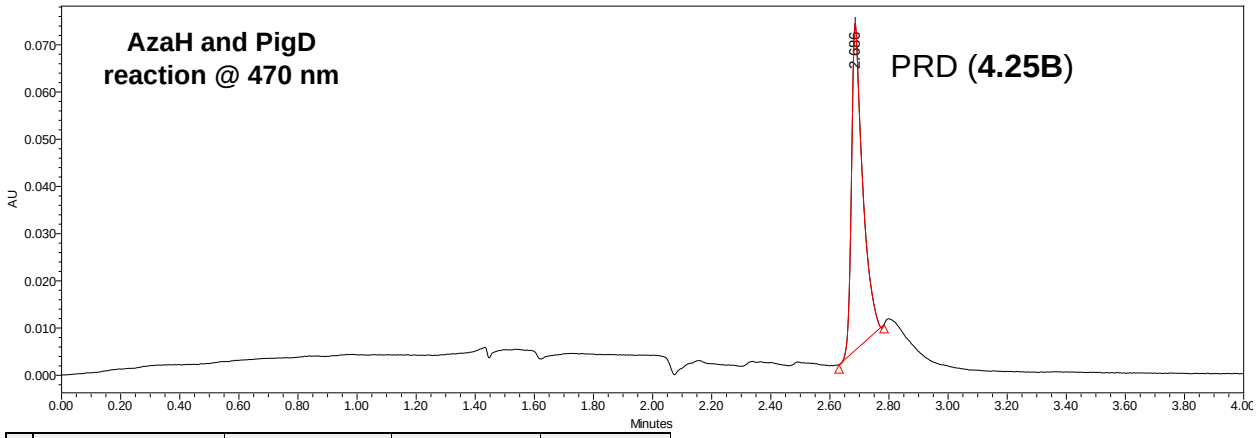
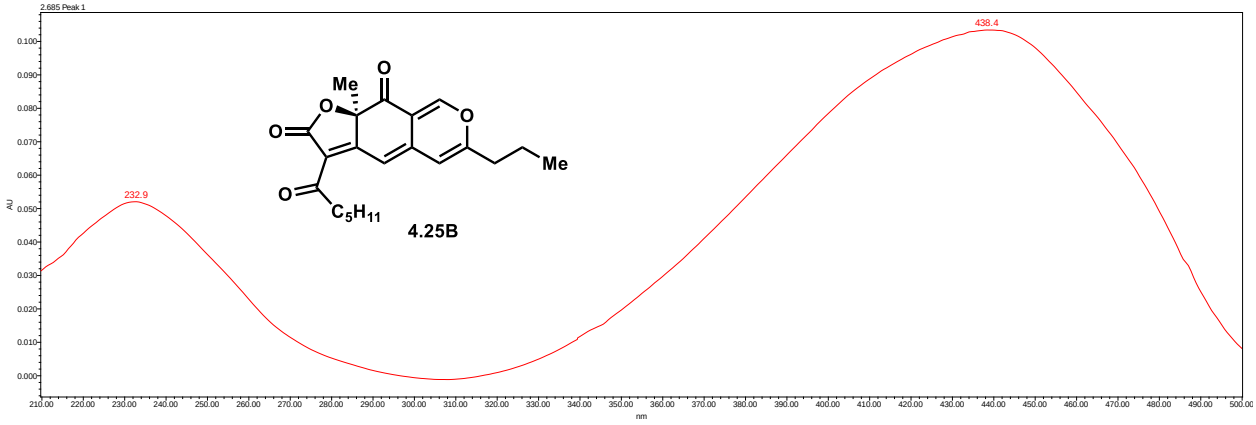

AzaH and PigD
EIC (391.1710-391.1770 m/z)
Exact Mass: 391.1751
+ H+
2.999
5254287.44
2.792
249644.18
AzaH only
EIC (391.1710-391.1770 m/z)
1.944
1006.88
Counts vs. Acquisition Time (min)
AzaH and PigD
ESI Scan (2.950-3.107 min)
373.1652
391.1759
AzaH only
ESI Scan (2.950-3.107 min)
279.0940
Counts vs. Mass-to-Charge (m/z)

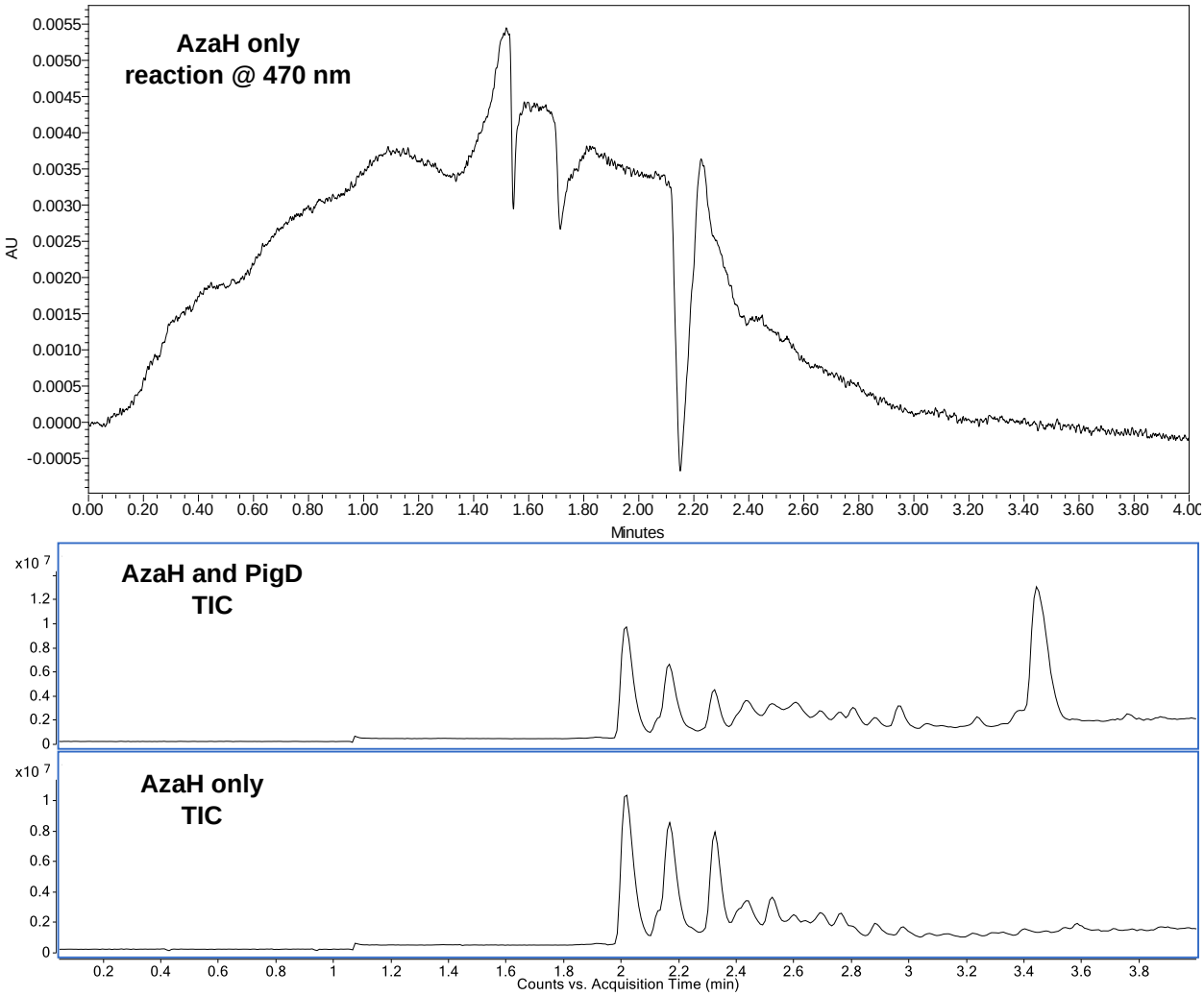
Figure 4.9. One-pot enzymatic transformation of **4.22B** by AzaH and PigD. PDA and LC-MS traces of enzymatic reaction and control reaction.





	Retention Time (min)	Area	% Area	Height
1	2.686	181862	100.00	69220




AzaH only
reaction @ 470 nm
AU
Minutes
AzaH and PigD
TIC
AzaH only
TIC
x10 7
Counts vs. Acquisition Time (min)

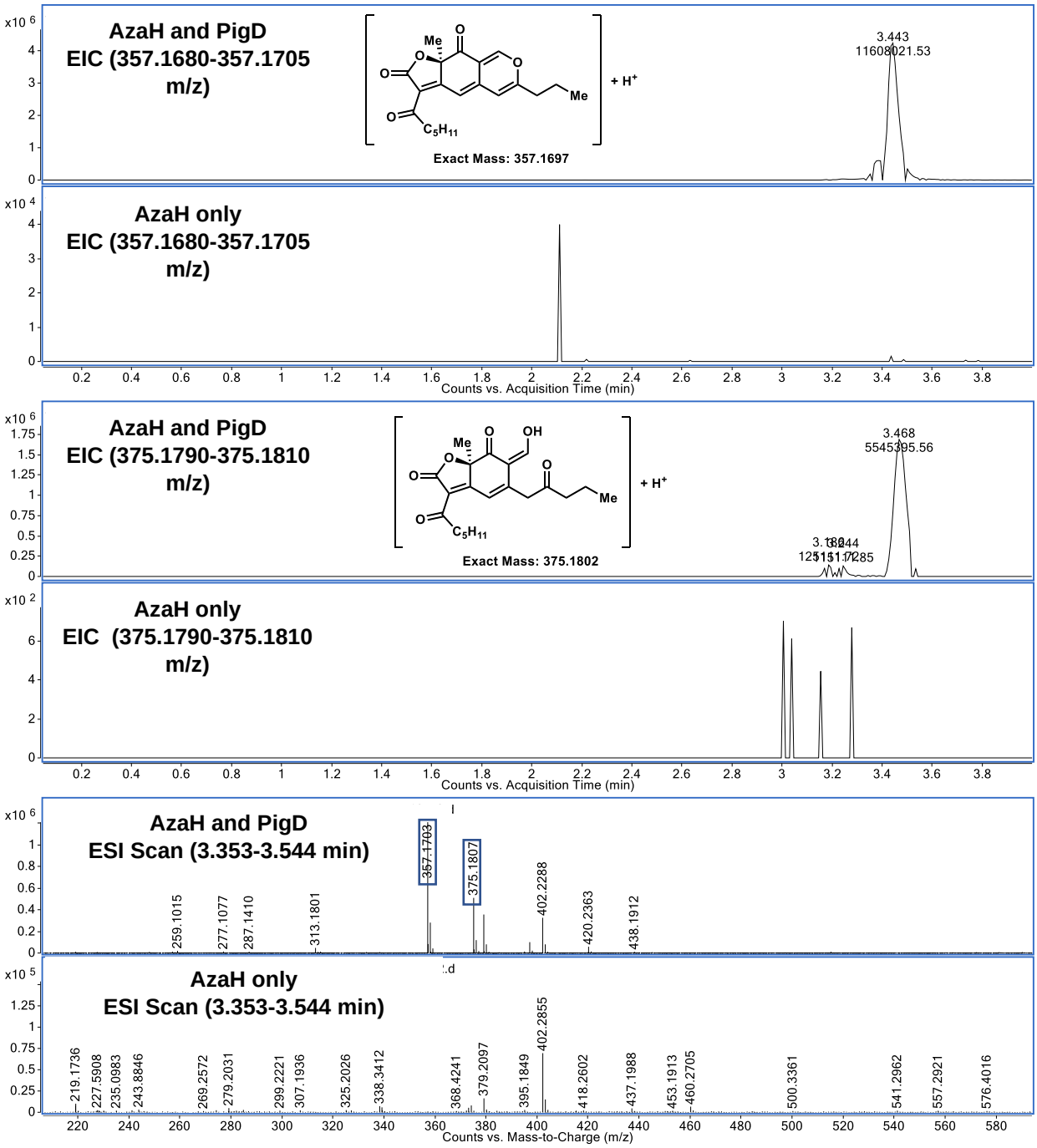
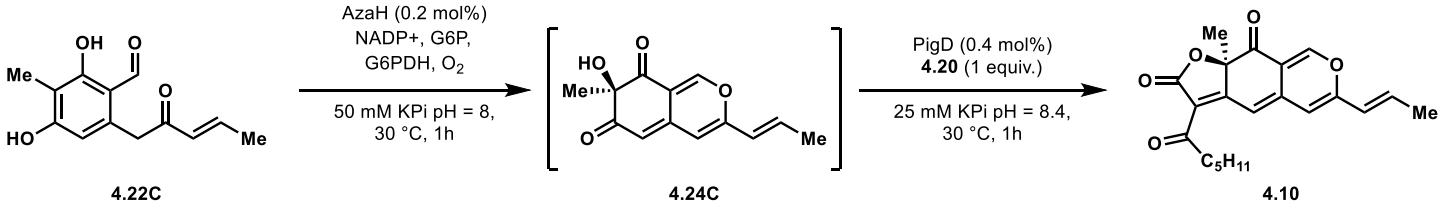
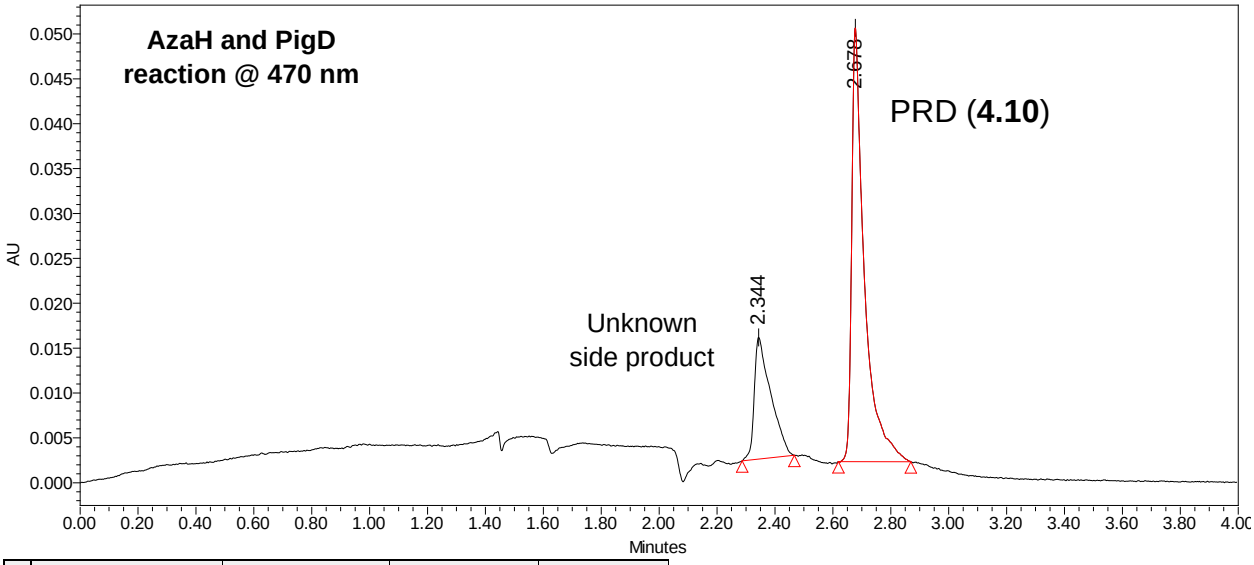
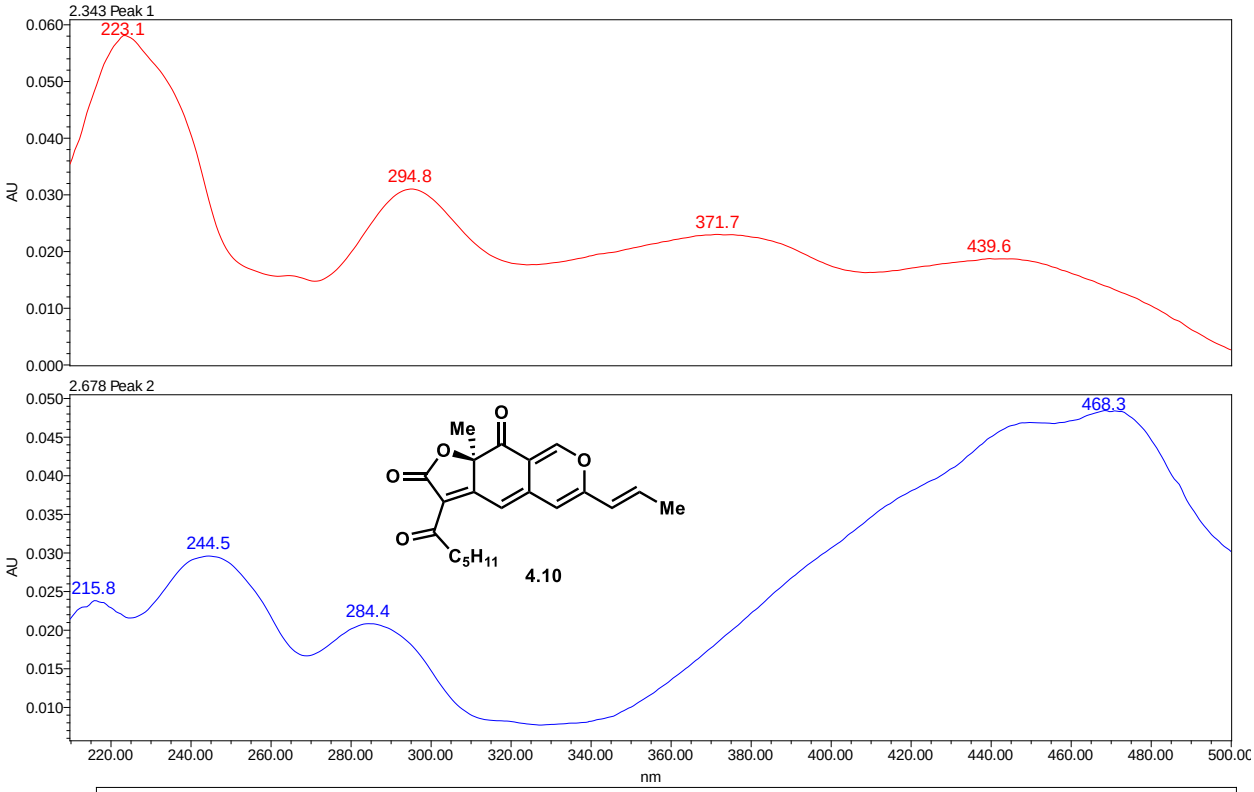

AzaH and PigD
EIC (357.1680-357.1705 m/z)
Exact Mass: 357.1697
3.443
11608021.53
AzaH only
EIC (357.1680-357.1705 m/z)
Counts vs. Acquisition Time (min)
AzaH and PigD
EIC (375.1790-375.1810 m/z)
Exact Mass: 375.1802
3.468
5545395.56
AzaH only
EIC (375.1790-375.1810 m/z)
Counts vs. Acquisition Time (min)
AzaH and PigD
ESI Scan (3.353-3.544 min)
357.1703
375.1807
259.1015
277.1077
287.1410
313.1801
402.2288
420.2363
438.1912
AzaH only
ESI Scan (3.353-3.544 min)
219.1736
227.5908
235.0983
243.8846
269.2572
279.2031
299.2221
307.1936
325.2026
338.3412
368.4241
379.2097
395.1849
402.2855
418.2602
437.1988
453.1913
460.2705
500.3361
541.2962
557.2921
576.4016
Counts vs. Mass-to-Charge (m/z)

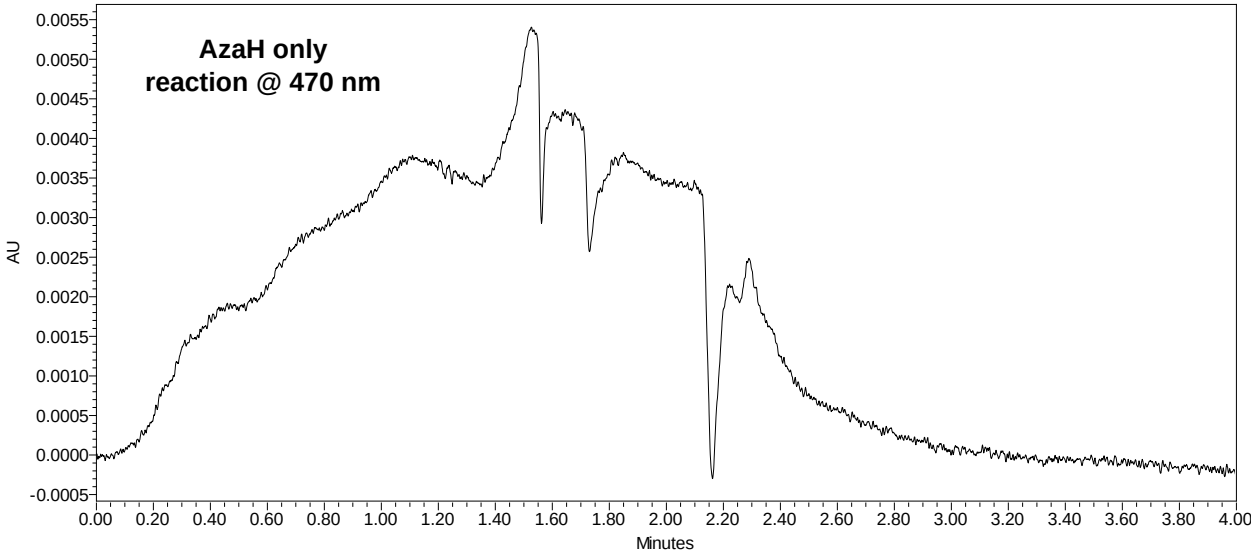
Figure 4.10. One-pot enzymatic transformation of **4.22C** by AzaH and PigD. PDA and LC-MS traces of enzymatic reaction and control reaction.

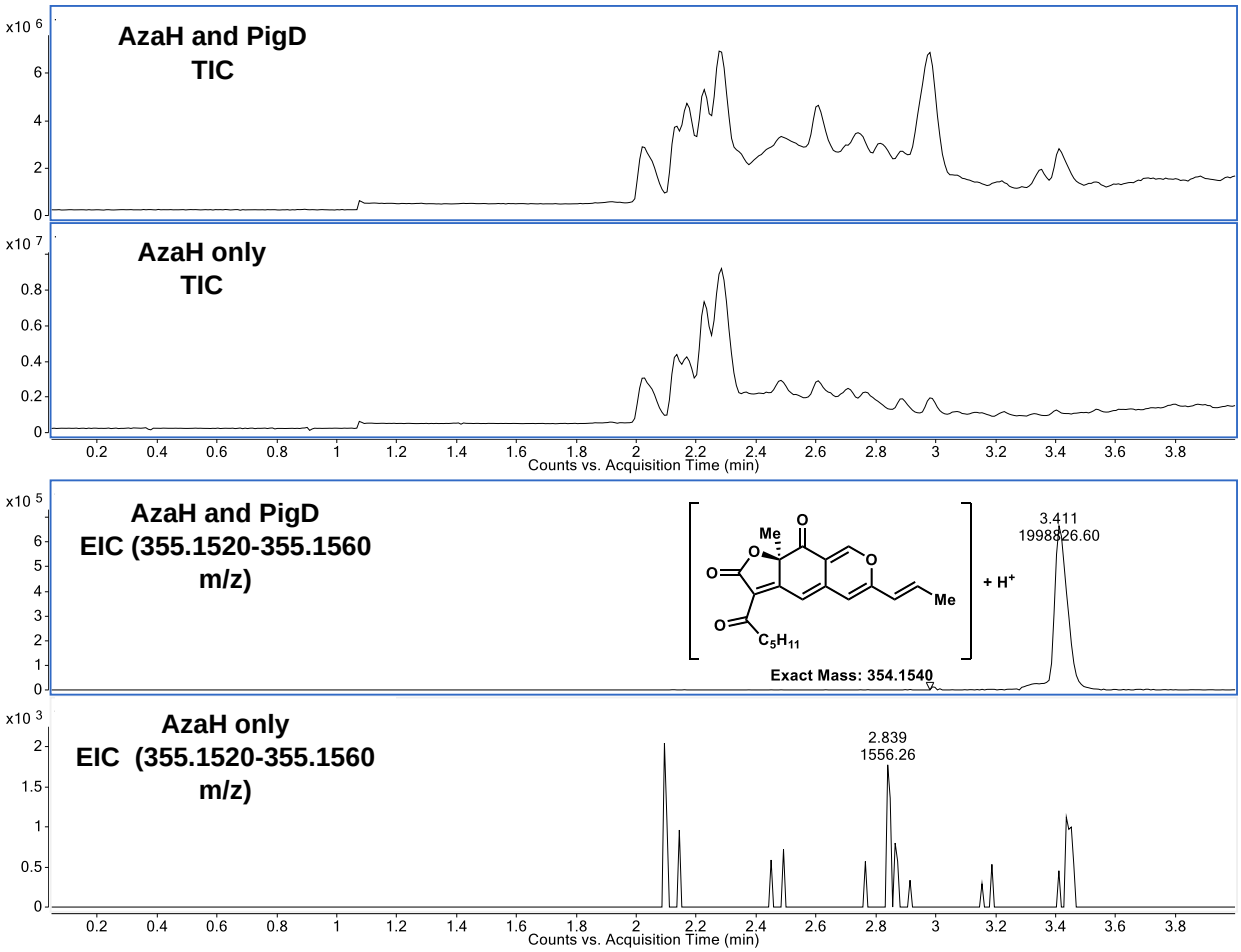




	Retention Time (min)	Area	% Area	Height
1	2.344	53466	26.65	13559
2	2.678	147134	73.35	48321


2.343 Peak 1
223.1
294.8
371.7
439.6
AU
2.678 Peak 2
468.3
244.5
215.8
284.4
Me
O
O
O
O
Me
C5H11
4.10
AU
nm


AzaH only
reaction @ 470 nm
AU
Minutes


AzaH and PigD
TIC
AzaH only
TIC
Counts vs. Acquisition Time (min)
AzaH and PigD
EIC (355.1520-355.1560
m/z)
Me
O
O
O
O
Me
O
C5H11
+ H+
Exact Mass: 354.1540
3.411
1998826.60
AzaH only
EIC (355.1520-355.1560
m/z)
2.839
1556.26
Counts vs. Acquisition Time (min)

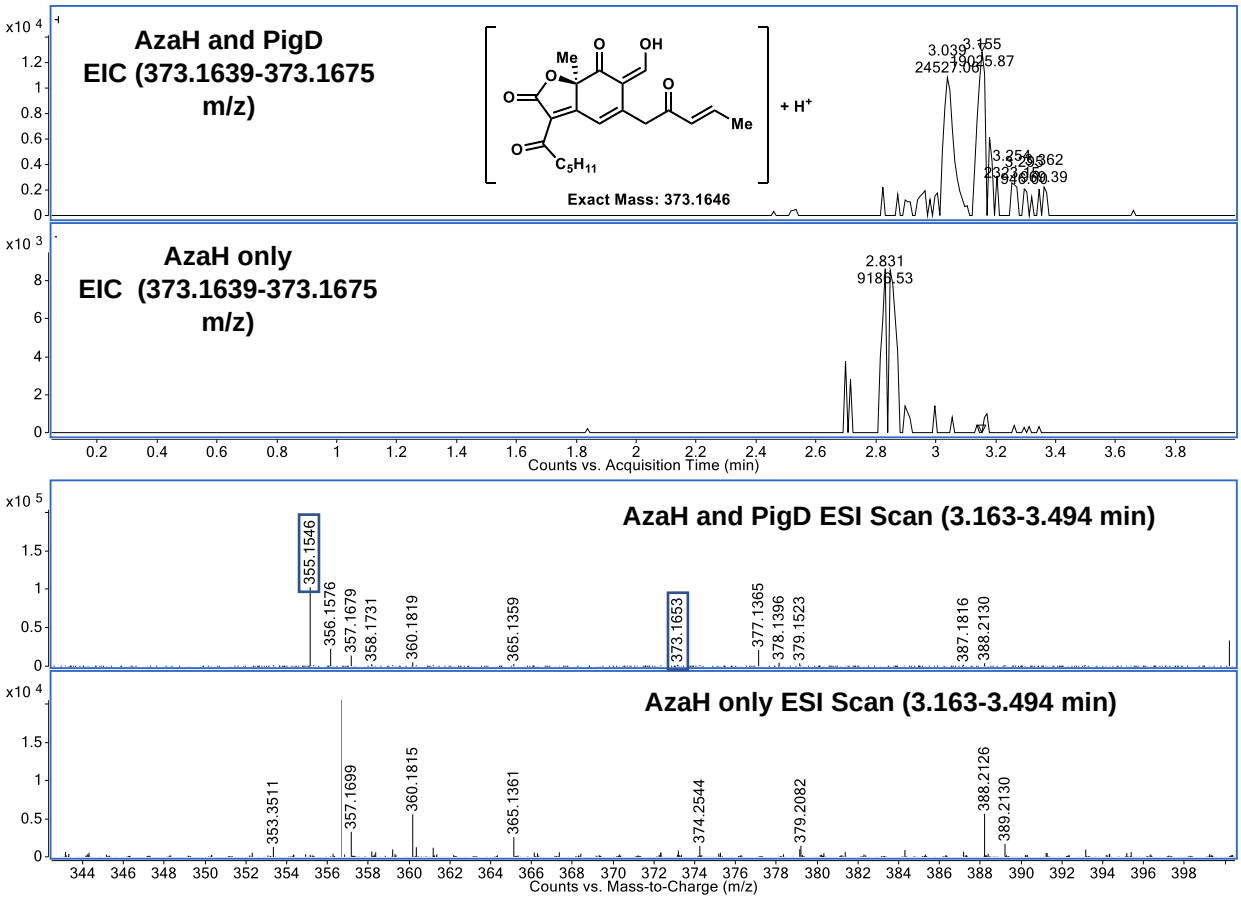
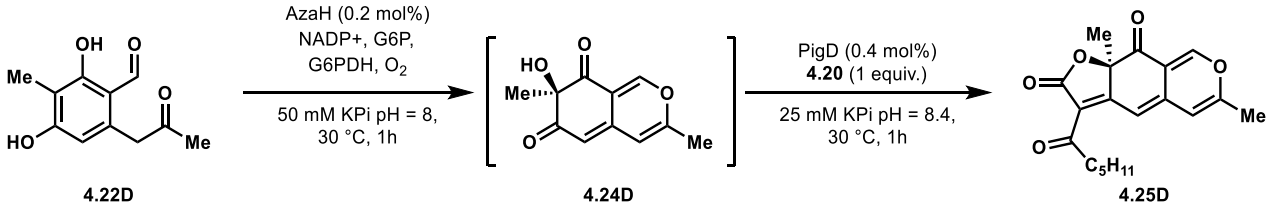
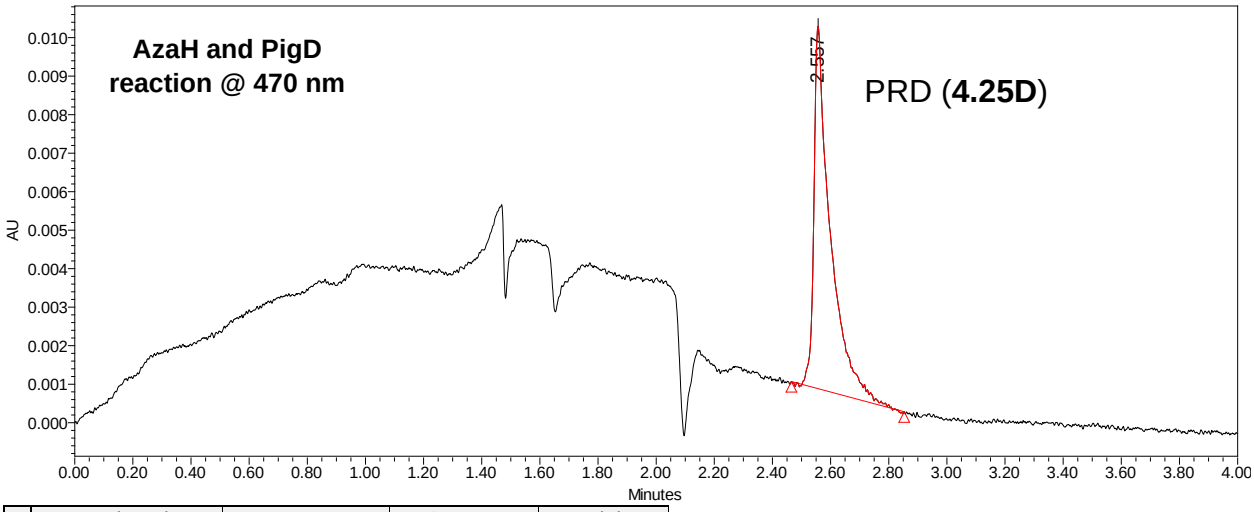
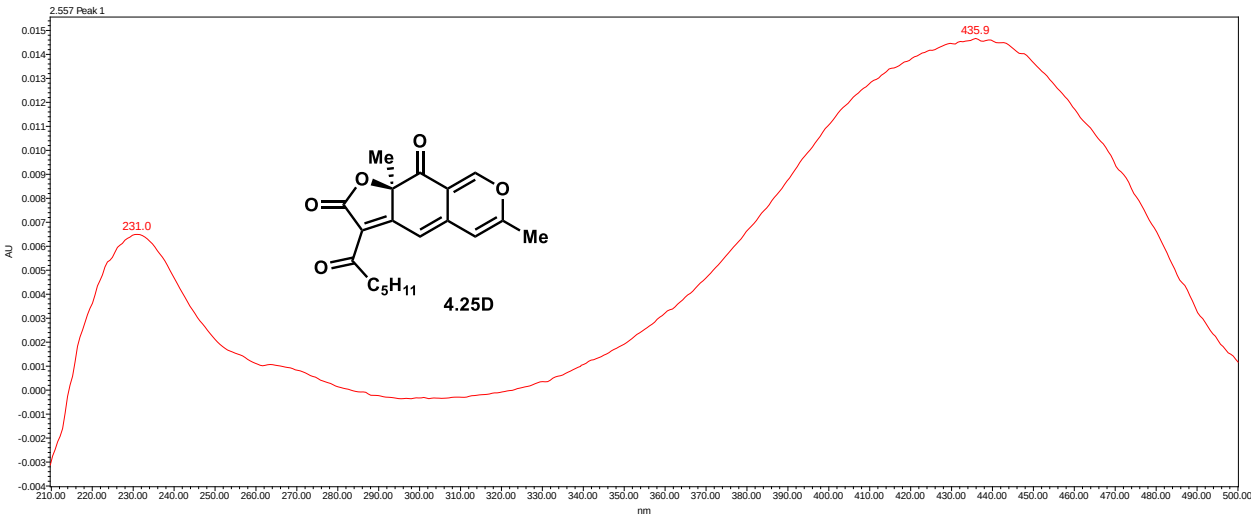

AzaH and PigD
EIC (373.1639-373.1675
m/z)
Me
O
OH
O
O
O
Me
O
C_5H_{11}
+ H^+
Exact Mass: 373.1646
3.039
24527.06
3.155
3.362
AzaH only
EIC (373.1639-373.1675
m/z)
2.831
9188.53
Counts vs. Acquisition Time (min)
AzaH and PigD ESI Scan (3.163-3.494 min)
355.1546
356.1576
357.1679
358.1731
360.1819
365.1359
373.1653
377.1365
378.1396
379.1523
387.1816
388.2130
AzaH only ESI Scan (3.163-3.494 min)
353.3511
357.1699
360.1815
365.1361
374.2544
379.2082
388.2126
389.2130
Counts vs. Mass-to-Charge (m/z)

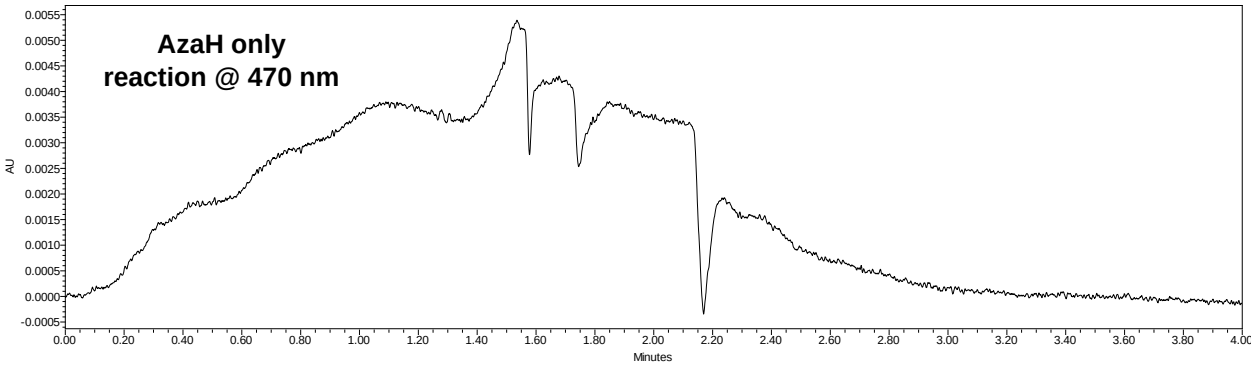
Figure 4.11. One-pot enzymatic transformation of **4.22D** by AzaH and PigD. PDA and LC-MS traces of enzymatic reaction and control reaction.

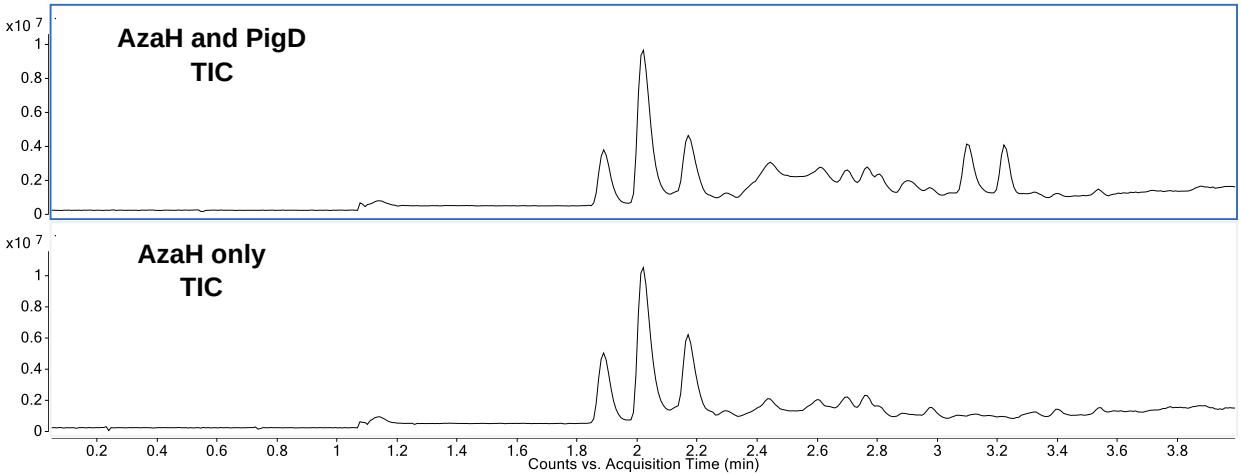


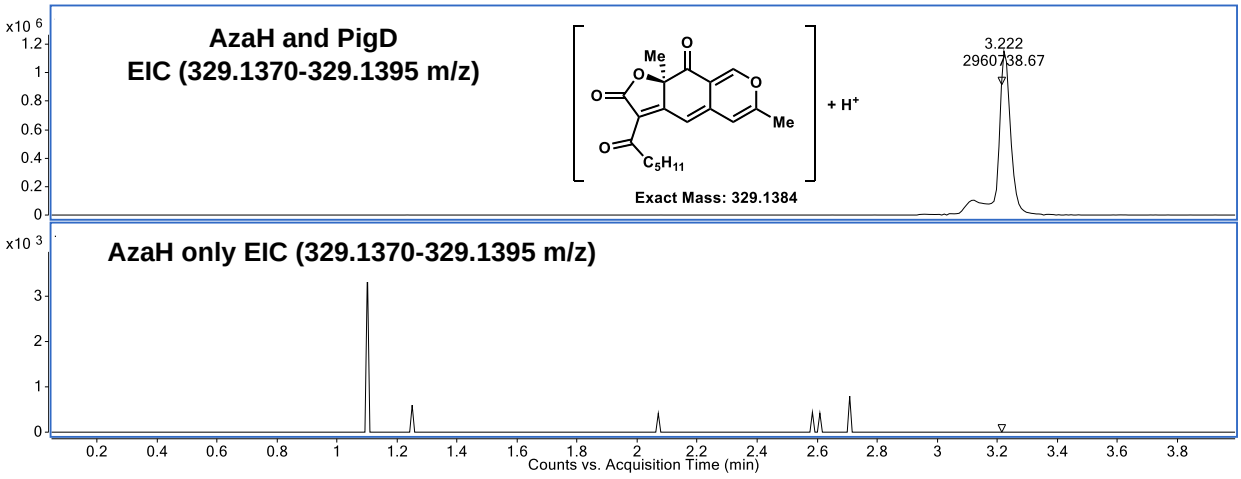


	Retention Time (min)	Area	% Area	Height
1	2.557	37052	100.00	9402




AzaH only
reaction @ 470 nm
AU
Minutes


AzaH and PigD
TIC
AzaH only
TIC
Counts vs. Acquisition Time (min)


AzaH and PigD
EIC (329.1370-329.1395 m/z)
Me
O
O
O
O
Me
O
C5H11
+ H+
Exact Mass: 329.1384
3.222
2960738.67
AzaH only EIC (329.1370-329.1395 m/z)
Counts vs. Acquisition Time (min)

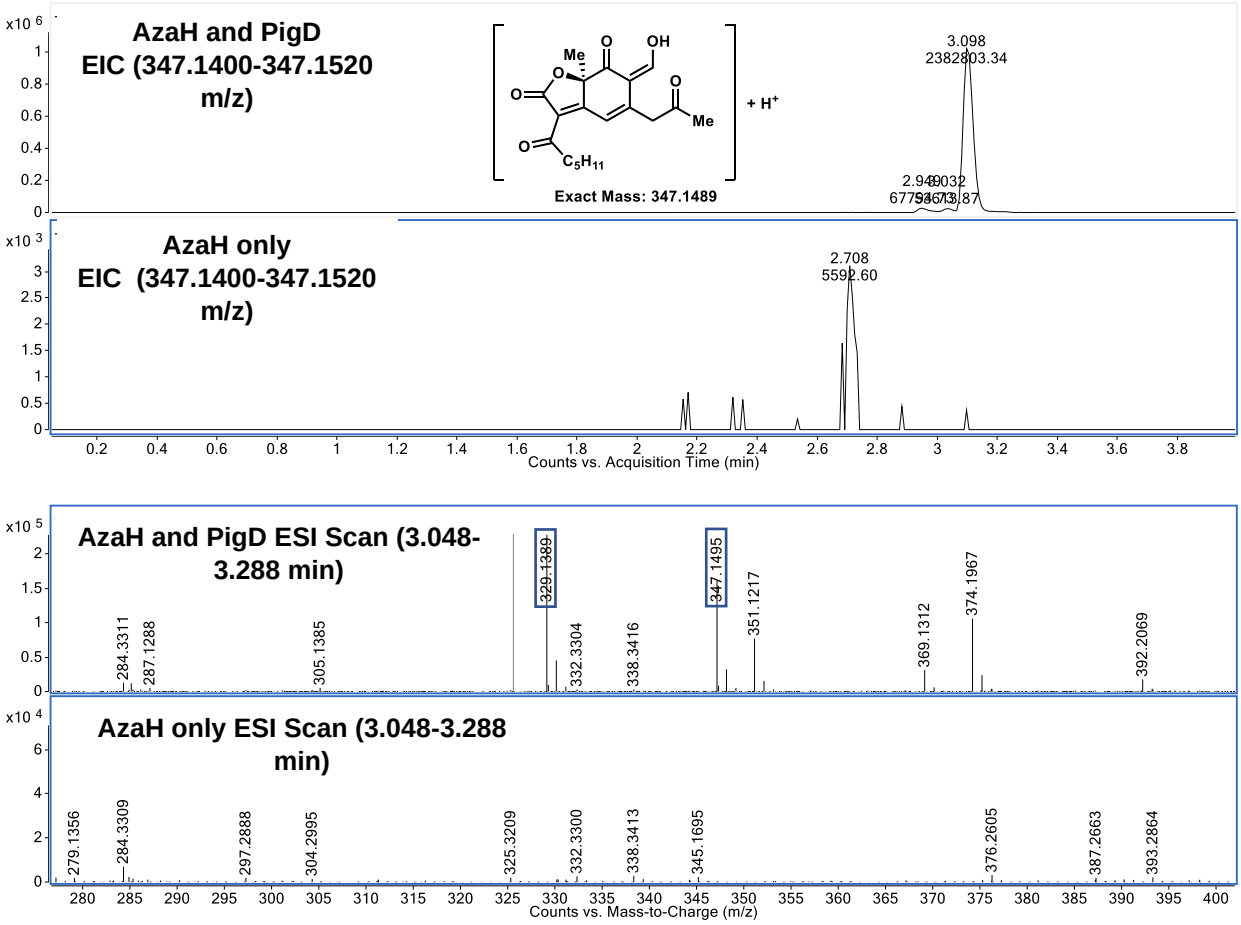
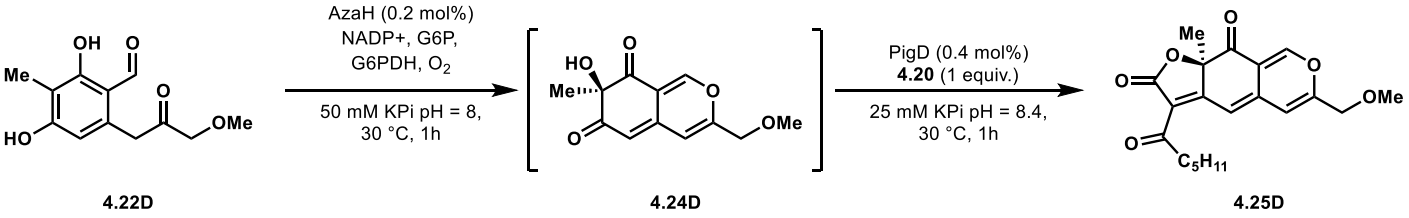
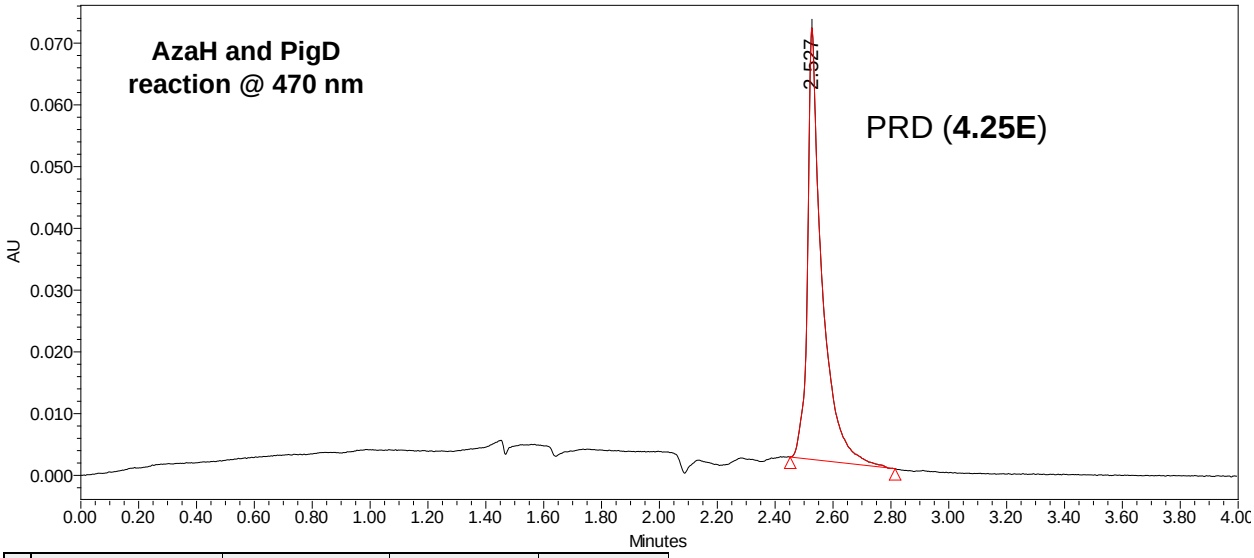
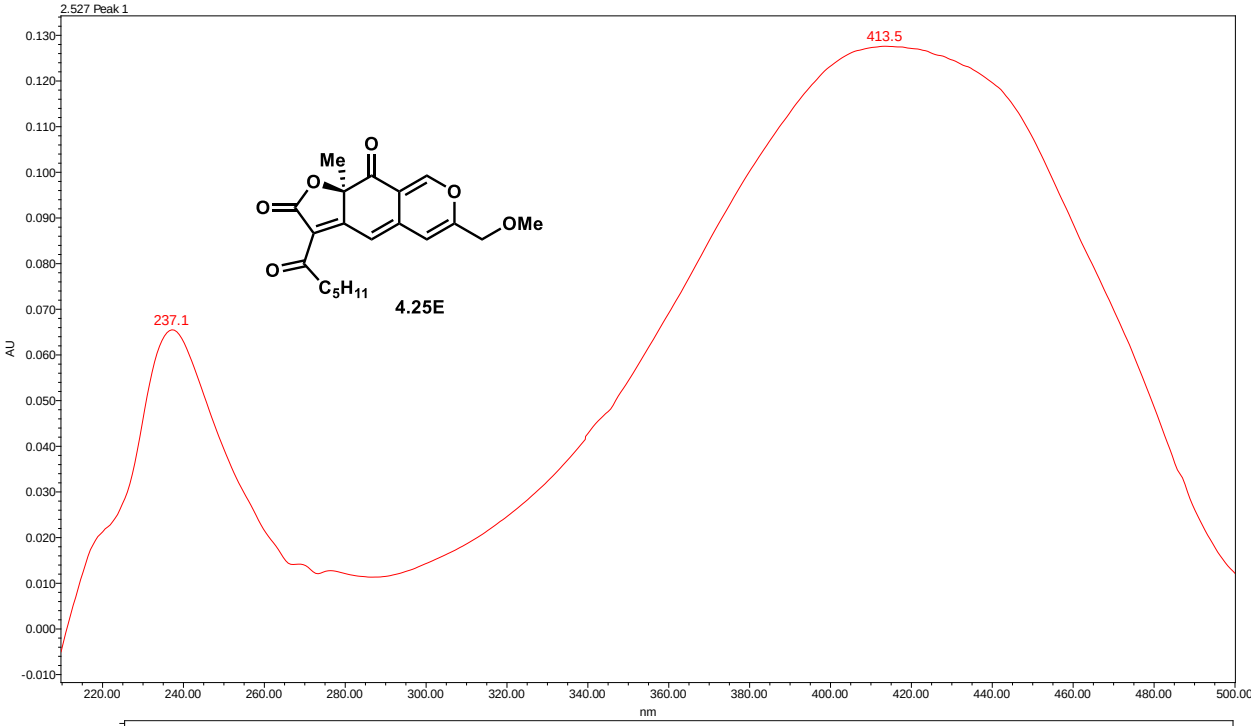

AzaH and PigD
EIC (347.1400-347.1520
m/z)
Me
O
OH
O
O
Me
O
C_5H_{11}
+ H^+
Exact Mass: 347.1489
3.098
2382803.34
AzaH only
EIC (347.1400-347.1520
m/z)
2.708
5592.60
Counts vs. Acquisition Time (min)
AzaH and PigD ESI Scan (3.048-
3.288 min)
329.1389
347.1495
284.3311
287.1288
305.1385
332.3304
338.3416
351.1217
369.1312
374.1967
392.2069
AzaH only ESI Scan (3.048-3.288
min)
279.1356
284.3309
297.2888
304.2995
325.3209
332.3300
338.3413
345.1695
376.2605
387.2663
393.2864
Counts vs. Mass-to-Charge (m/z)

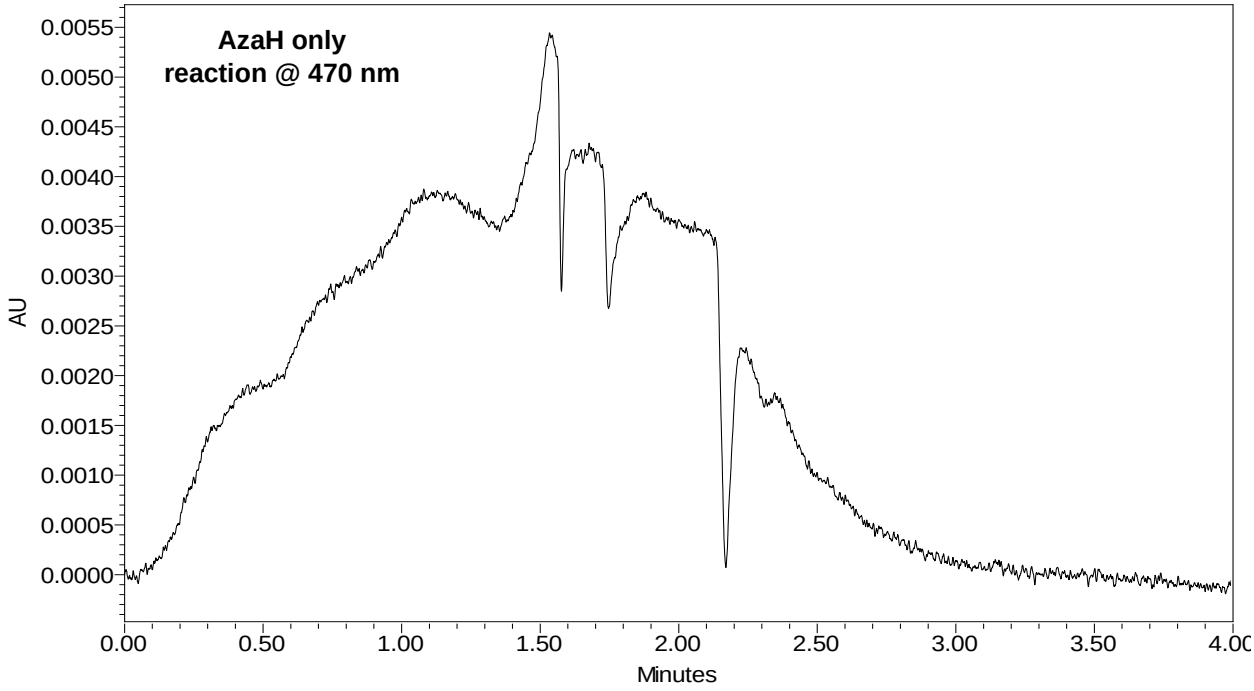
Figure 4.12. One-pot enzymatic transformation of **4.22E** by AzaH and PigD. PDA and LC-MS traces of enzymatic reaction and control reaction.

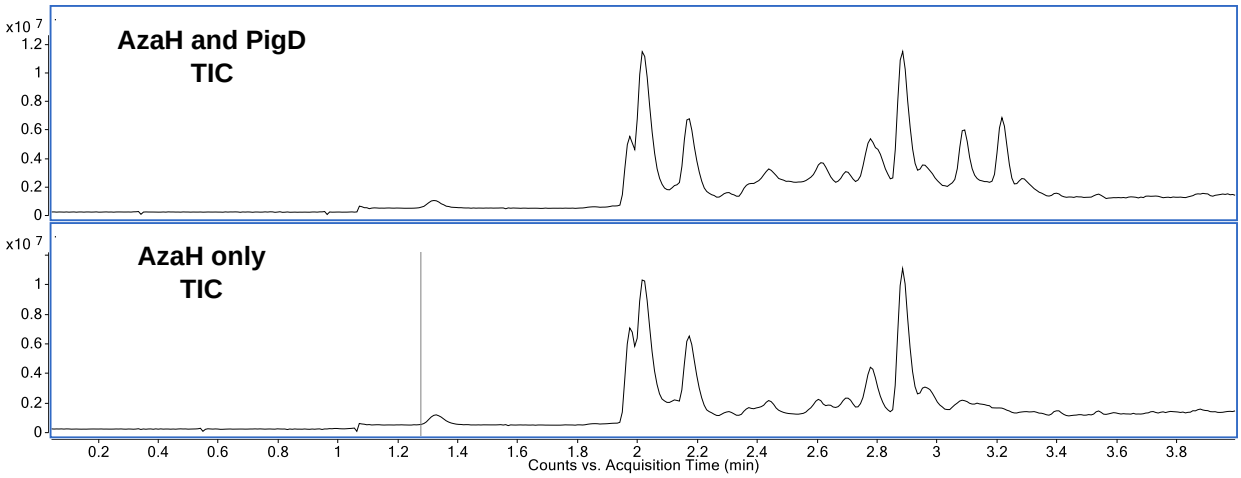


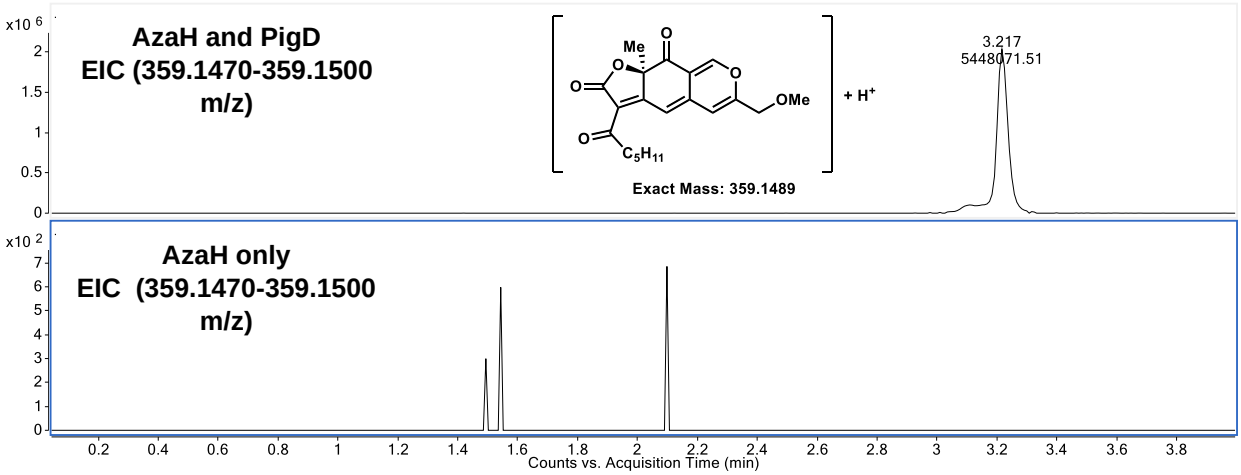


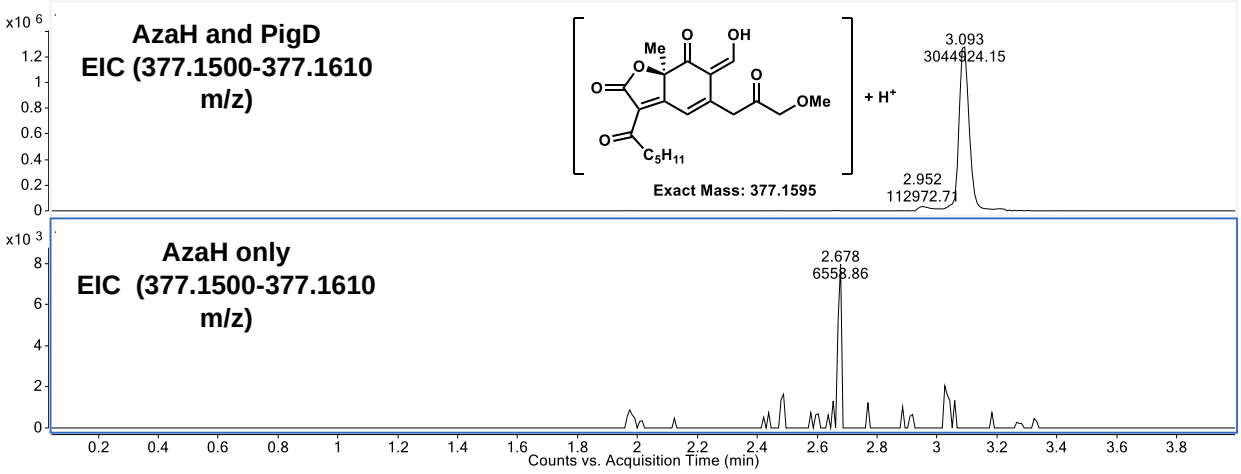
	Retention Time (min)	Area	% Area	Height
1	2.527	241845	100.00	69893

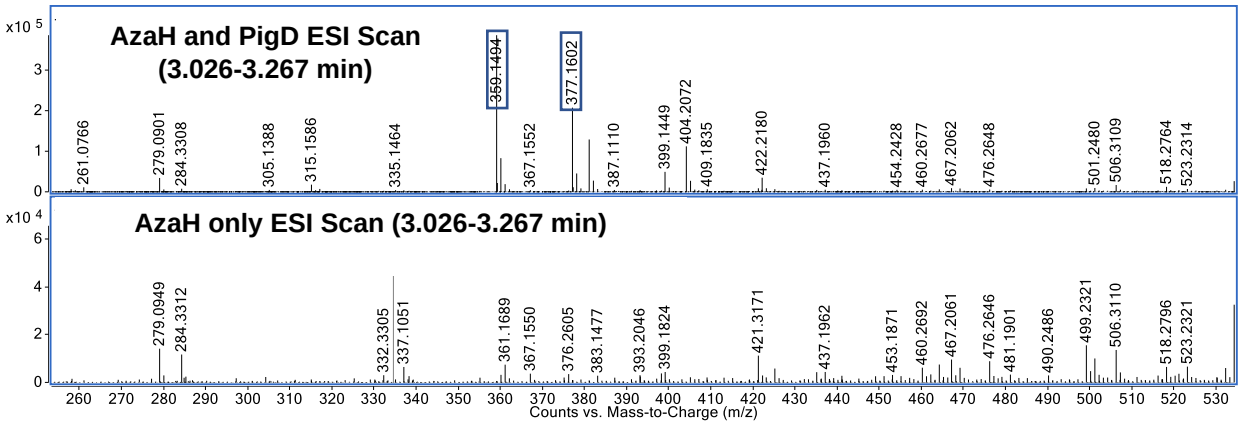

2.527 Peak 1
413.5
237.1
Me
O
O
O
O
OMe
O
C_5H_{11}
4.25E
AU
nm


AzaH only
reaction @ 470 nm
AU
Minutes

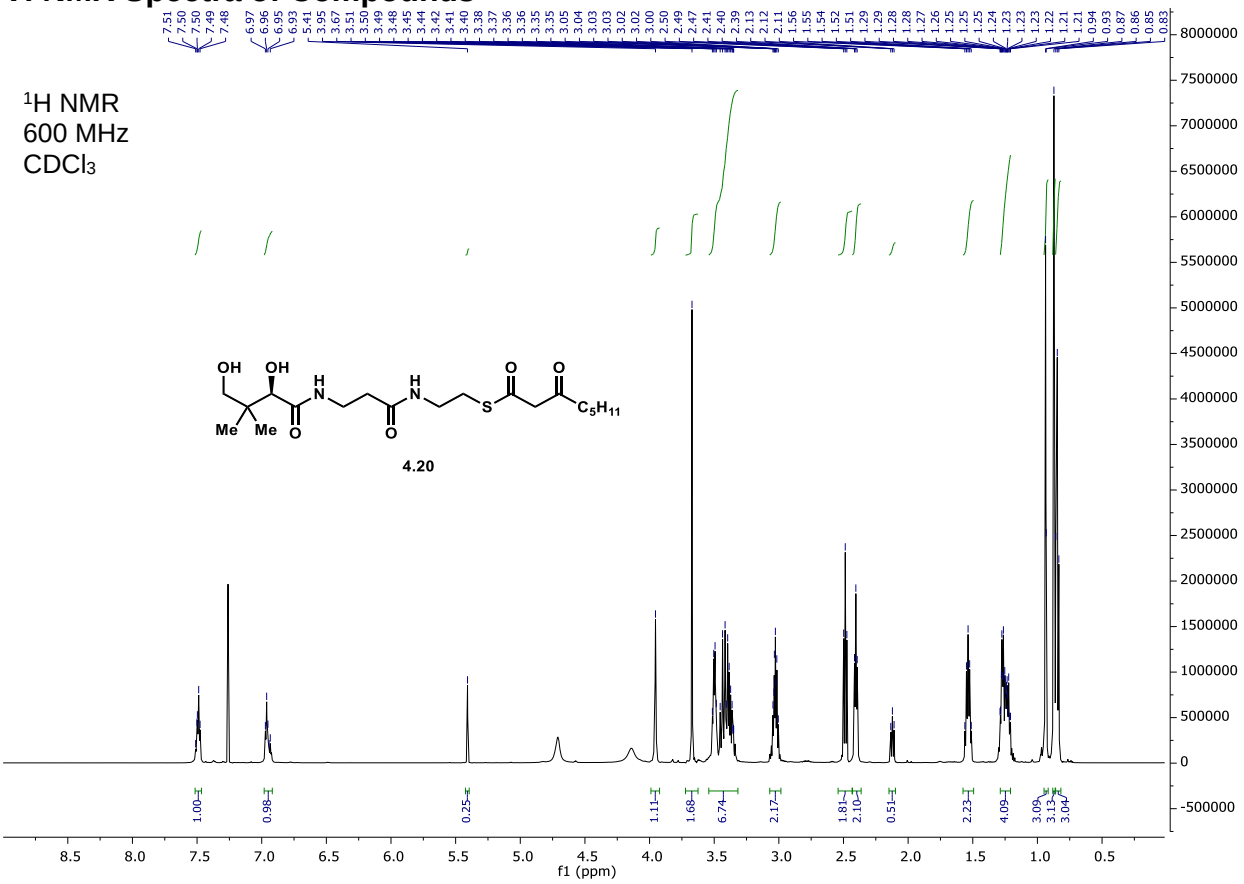

AzaH and PigD
TIC
AzaH only
TIC
Counts vs. Acquisition Time (min)

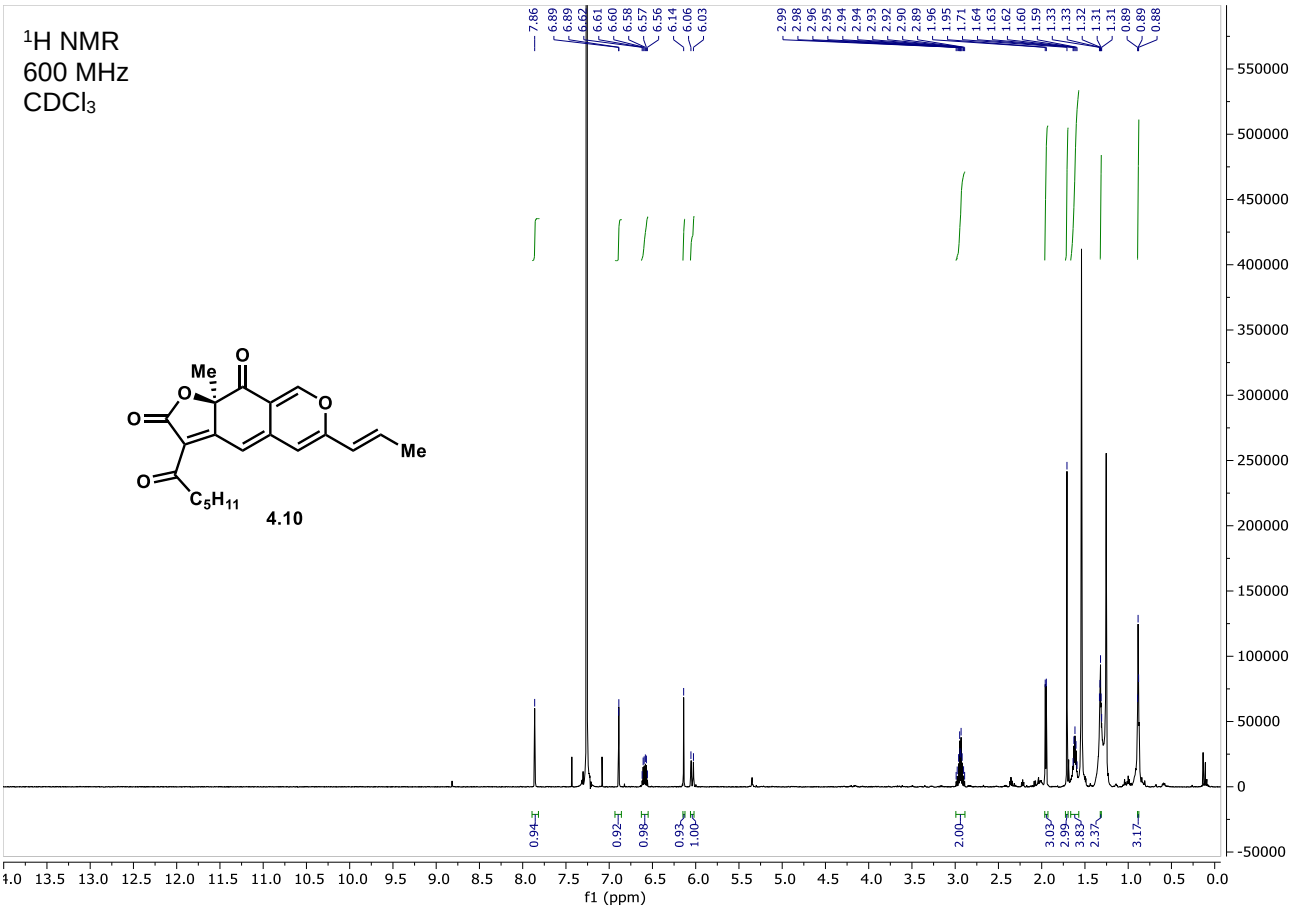

AzaH and PigD
EIC (359.1470-359.1500
m/z)
Me
O
OMe
+ H+
C5H11
Exact Mass: 359.1489
3.217
5448071.51
AzaH only
EIC (359.1470-359.1500
m/z)
Counts vs. Acquisition Time (min)

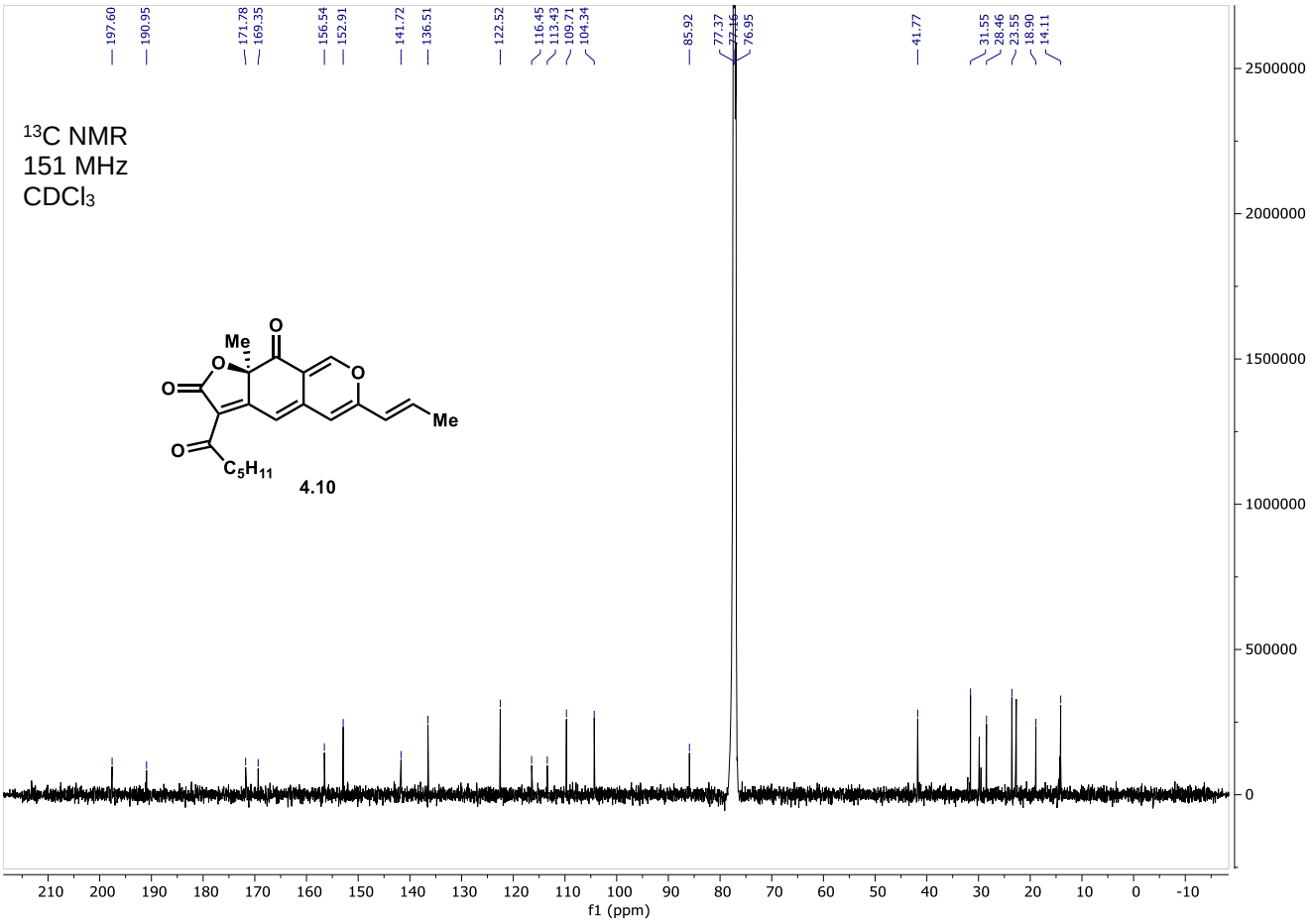

AzaH and PigD
EIC (377.1500-377.1610
m/z)
Me
O
OH
OMe
+ H+
C5H11
Exact Mass: 377.1595
3.093
3044824.15
2.952
112972.71
AzaH only
EIC (377.1500-377.1610
m/z)
2.678
6558.86
Counts vs. Acquisition Time (min)

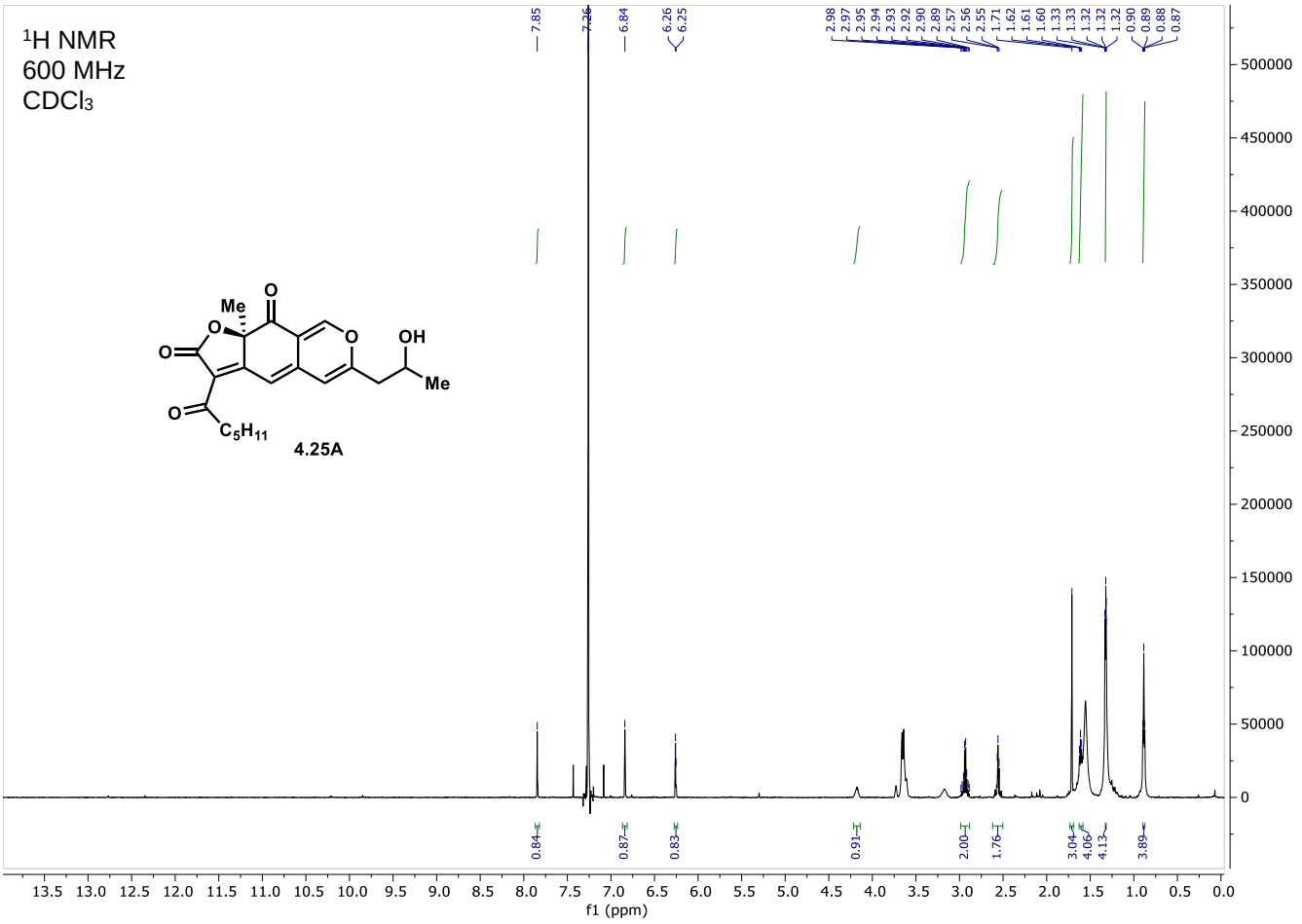

AzaH and PigD ESI Scan
(3.026-3.267 min)
x10 5
3
2
1
0
261.0766
279.0901
284.3308
305.1388
315.1586
335.1464
369.1494
367.1552
377.1602
387.1110
399.1449
404.2072
409.1835
422.2180
437.1960
454.2428
460.2677
467.2062
476.2648
501.2480
506.3109
518.2764
523.2314
AzaH only ESI Scan (3.026-3.267 min)
x10 4
6
4
2
0
279.0949
284.3312
332.3305
337.1051
361.1689
367.1550
376.2605
383.1477
393.2046
399.1824
421.3171
437.1962
453.1871
460.2692
467.2061
476.2646
481.1901
490.2486
499.2321
506.3110
518.2796
523.2321
260 270 280 290 300 310 320 330 340 350 360 370 380 390 400 410 420 430 440 450 460 470 480 490 500 510 520 530
Counts vs. Mass-to-Charge (m/z)

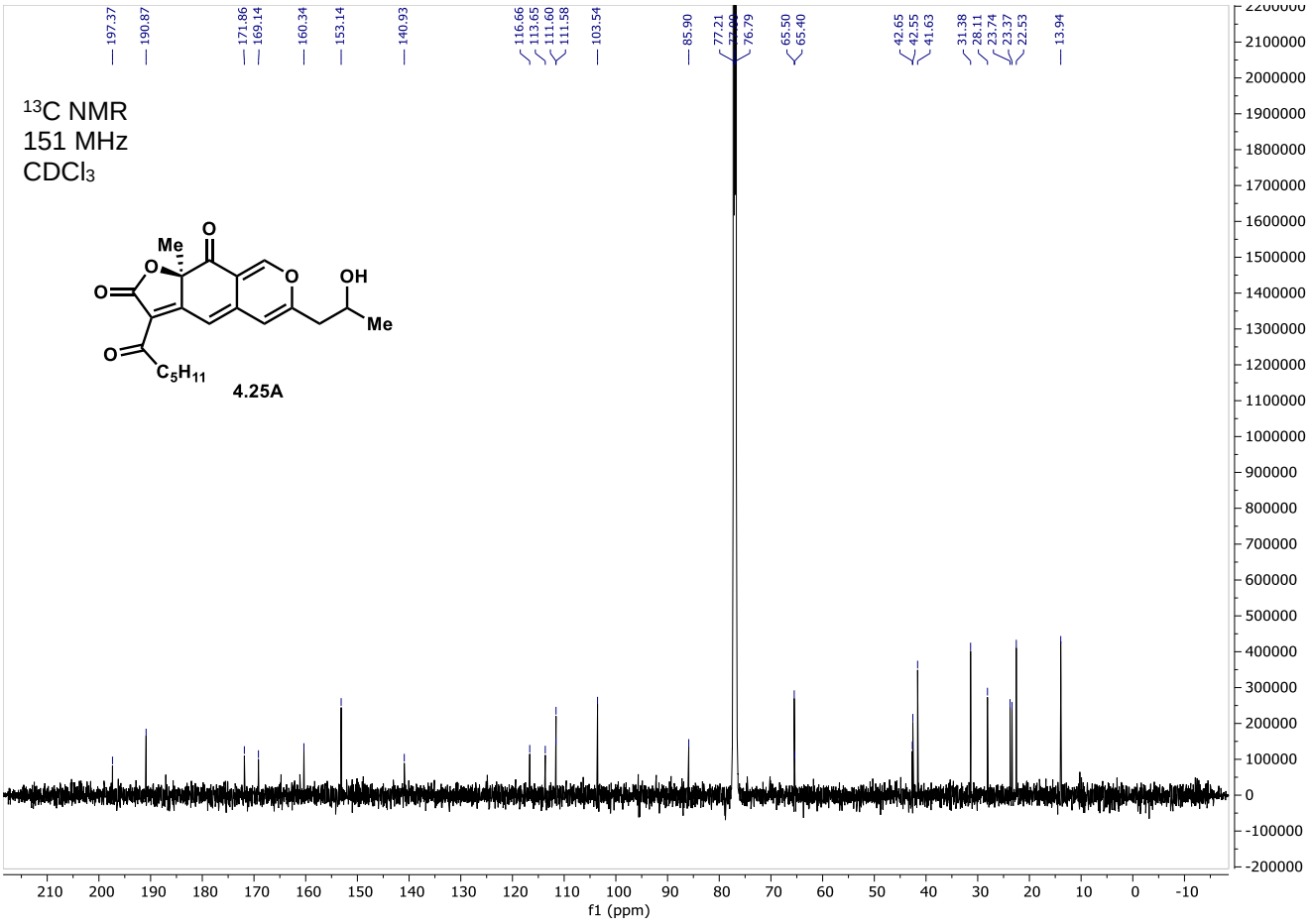
V. NMR Spectra of Compounds

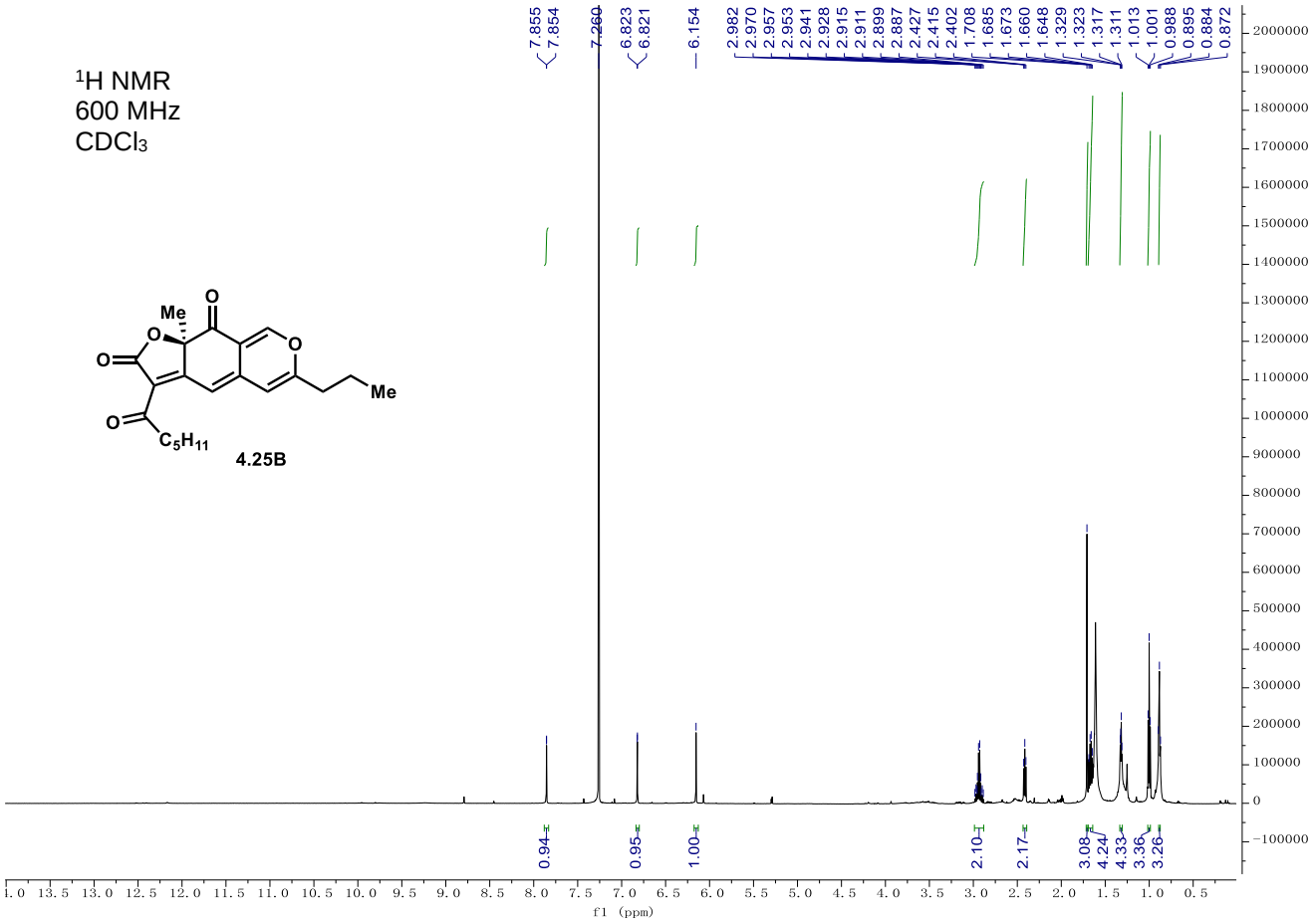


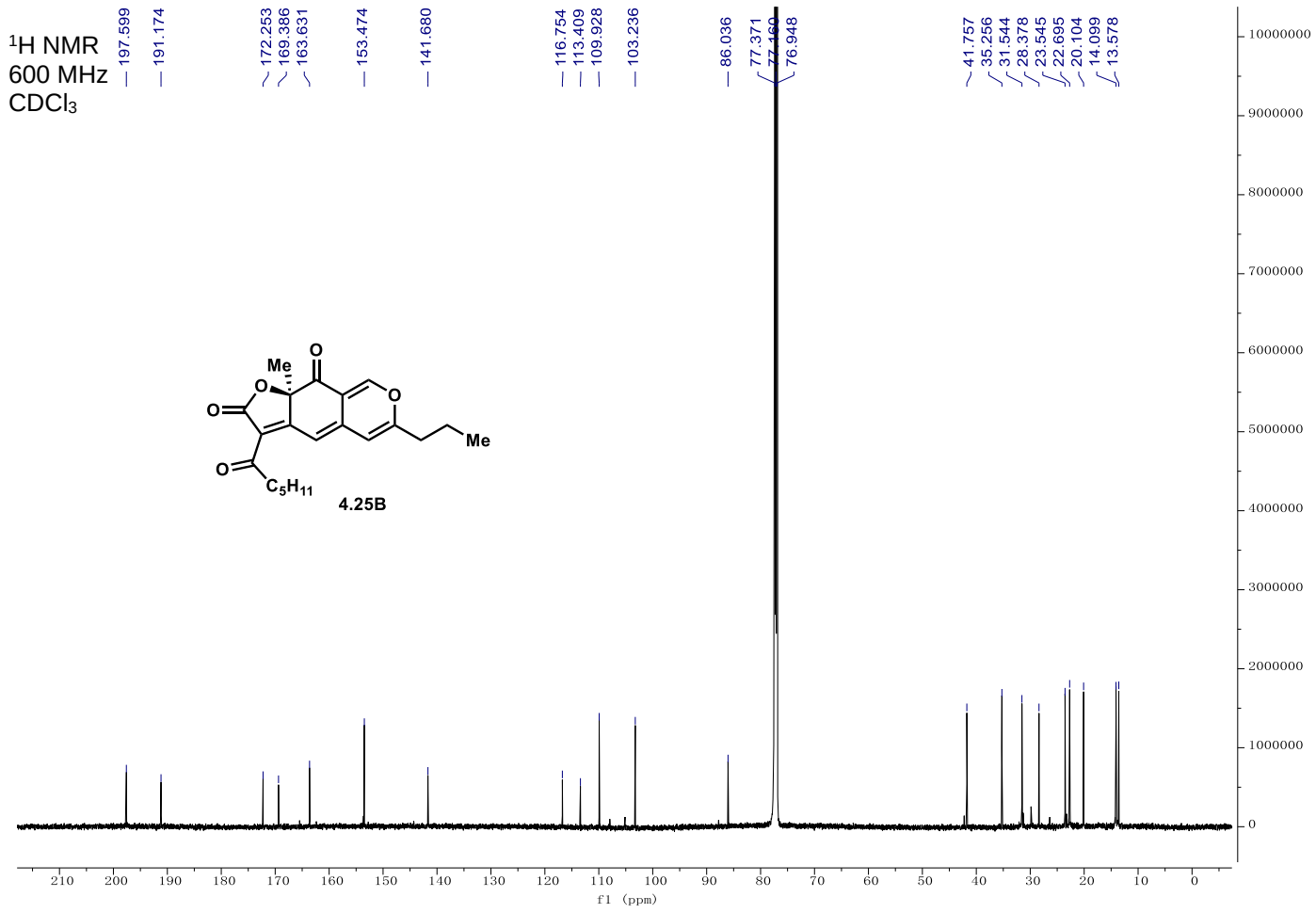

1H NMR
600 MHz
CDCl3
Me
O
O
O
O
O
Me
C5H11
4.10
f1 (ppm)


197.60
190.95
171.78
169.35
156.54
152.91
141.72
136.51
122.52
116.45
113.43
109.71
104.34
85.92
77.37
77.16
76.95
41.77
31.55
28.46
23.55
18.90
14.11
13C NMR
151 MHz
CDCl3
Me
O
O
O
O
Me
O
C5H11
4.10
2500000
2000000
1500000
1000000
500000
0
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)


1H NMR
600 MHz
CDCl3
Me
O
O
O
O
OH
Me
O
C5H11
4.25A
7.85
7.26
6.84
6.26
6.25
2.98
2.97
2.95
2.94
2.93
2.92
2.90
2.89
2.57
2.56
2.55
1.71
1.62
1.61
1.60
1.33
1.32
1.32
1.32
0.90
0.89
0.88
0.87
0.84
0.87
0.83
0.91
2.00
1.76
3.04
4.06
4.13
3.89
f1 (ppm)


197.37
190.87
171.86
169.14
160.34
153.14
140.93
116.66
113.65
111.60
111.58
103.54
85.90
77.21
76.79
65.50
65.40
42.65
42.55
41.63
31.38
28.11
23.74
23.37
22.53
13.94
13C NMR
151 MHz
CDCl3
Me
O
O
O
O
OH
Me
O
C5H11
4.25A
f1 (ppm)


1H NMR
600 MHz
CDCl3
Me
O
O
O
O
O
Me
C5H11
4.25B
7.855
7.854
7.260
6.823
6.821
6.154
2.982
2.970
2.957
2.953
2.941
2.928
2.915
2.911
2.899
2.887
2.427
2.415
2.402
1.708
1.685
1.673
1.660
1.648
1.329
1.323
1.317
1.311
1.013
1.001
0.988
0.895
0.884
0.872
0.94
0.95
1.00
2.10
2.17
3.08
4.24
4.33
3.36
3.26
f1 (ppm)



4.6 References

(1) Pavesi, C.; Flon, V.; Mann, S.; Leleu, S.; Prado, S.; Franck, X. Biosynthesis of azaphilones: a review. *Nat. Prod. Rep.* **2021**.

(2) Fisch, K. M. Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS. *RSC Adv.* **2013**, *3*, 18228-18247.

(3) Wang, J.; Zhang, R.; Chen, X.; Sun, X.; Yan, Y.; Shen, X.; Yuan, Q. Biosynthesis of aromatic polyketides in microorganisms using type II polyketide synthases. *Microb. Cell Fact.* **2020**, *19*, 110.

(4) Campbell, C. D.; Vederas, J. C. Biosynthesis of lovastatin and related metabolites formed by fungal iterative PKS enzymes. *Biopolymers* **2010**, *93*, 755-63.

(5) Nivina, A.; Yuet, K. P.; Hsu, J.; Khosla, C. Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chemica. Rev.* **2019**, *119*, 12524-12547.

(6) Gallimore, A. R.; Spencer, J. B. Stereochemical Uniformity in Marine Polyether Ladders—Implications for the Biosynthesis and Structure of Maitotoxin. *Angew. Chem. Int. Ed.* **2006**, *45*, 4406-4413.

(7) Xue, Y.; Sherman, D. H. Biosynthesis and Combinatorial Biosynthesis of Pikromycin-Related Macrolides in Streptomyces venezuelae. *Metab. Eng.* **2001**, *3*, 15-26.

(8) Huffman, M. A.; Fryszkowska, A.; Alvizo, O.; Borra-Garske, M.; Campos, K. R.; Canada, K. A.; Devine, P. N.; Duan, D.; Forstater, J. H.; Grosser, S. T.; Halsey, H. M.; Hughes, G. J.; Jo, J.; Joyce, L. A.; Kolev, J. N.; Liang, J.; Maloney, K. M.; Mann, B. F.; Marshall, N. M.; McLaughlin, M.; Moore, J. C.; Murphy, G. S.; Nawrat, C. C.; Nazor, J.; Novick, S.; Patel, N. R.; Rodriguez-Granillo, A.; Robaire, S. A.; Sherer, E. C.; Truppo, M. D.; Whittaker, A. M.; Verma, D.; Xiao, L.; Xu, Y.; Yang, H. Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science* **2019**, *366*, 1255-1259.
(9) Sheldon, R. A.; Woodley, J. M. Role of Biocatalysis in Sustainable Chemistry. *Chem. Rev.* **2018**, *118*, 801–838.
(10) Winkler, C. K.; Schrittwieser, J. H.; Kroutil, W. Power of Biocatalysis for Organic Synthesis. *ACS Cen. Sci.* **2021**.
(11) Hughes, G.; Lewis, J. C. Introduction: Biocatalysis in Industry. *Chem. Rev.* **2018**, *118*, 1-3.
(12) Wandrey, C.; Liese, A.; Kihumbu, D. Industrial Biocatalysis: Past, Present, and Future. *Org. Process Res. Dev.* **2000**, *4*, 286-290.
(13) Bornscheuer, U. T.; Buchholz, K. Highlights in Biocatalysis – Historical Landmarks and Current Trends. *Eng.Life Sci.* **2005**, *5*, 309-323.
(14) Rodríguez Benítez, A.; Narayan, A. R. H. Frontiers in Biocatalysis: Profiling Function across Sequence Space. *ACS Cent. Sci.* **2019**, *5*, 1747-1749.
(15) Devine, P. N.; Howard, R. M.; Kumar, R.; Thompson, M. P.; Truppo, M. D.; Turner, N. J. Extending the application of biocatalysis to meet the challenges of drug development. *Nat. Rev. Chem.* **2018**, *2*, 409-421.
(16) Truppo, M. D. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med. Chem. Lett.* **2017**, *8*, 476-480.
(17) Reetz, M. T. Biocatalysis in Organic Chemistry and Biotechnology: Past, Present, and Future. *J. Am. Chem. Soc.* **2013**, *135*, 12480-12496.
(18) Hughes, D. L. Biocatalysis in Drug Development—Highlights of the Recent Patent Literature. *Org. Process Res. Dev.* **2018**, *22*, 1063-1080.
(19) Heckmann, C. M.; Paradisi, F. Looking Back: A Short History of the Discovery of Enzymes and How They Became Powerful Chemical Tools. *Chem. Cat. Chem.* **2020**, *12*, 6082-6102.
(20) Hughes, R. A.; Ellington, A. D. Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. *Cold Spring Harb. Perspect. Biol.* **2017**, *9*.
(21) Siloto, R. M. P.; Weselake, R. J. Site saturation mutagenesis: Methods and applications in protein engineering. *Biocatal. Agric. Biotechnol.* **2012**, *1*, 181-189.
(22) Wilding, M.; Scott, C.; Warden, A. C. Computer-Guided Surface Engineering for Enzyme Improvement. *Sci. Rep.* **2018**, *8*, 11998.
(23) Pengyi, Y.; Yee Hwa, Y.; Bing, B. Z.; Albert, Y. Z. A Review of Ensemble Methods in Bioinformatics. *Curr. Bioinform.* **2010**, *5*, 296-308.
(24) Roumpeka, D. D.; Wallace, R. J.; Escalettes, F.; Fotheringham, I.; Watson, M. A Review of Bioinformatics Tools for Bio-Prospecting from Metagenomic Sequence Data. *Front. Genet.* **2017**, *8*.
(25) Chen, R.; Liu, Q.; He, Y.; He, K.; Ding, X.; Kang, L.; Guo, X.; Xie, N.; Zhou, Y.; Lu, Y.; Cox, R. J.; Molnár, I.; Li, M.; Shao, Y.; Chen, F. Orange, red, yellow:

biosynthesis of azaphilone pigments in Monascus fungi. *Chem. Sci.* **2017**, *8*, 4917-4925.
(26) Huang, X.; Zhang, W.; Tang, S.; Wei, S.; Lu, X. Collaborative Biosynthesis of a Class of Bioactive Azaphilones by Two Separate Gene Clusters Containing Four PKS/NRPSs with Transcriptional Crosstalk in Fungi. *Angew. Chem. Int. Ed.* **2020**, *59*, 4349-4353.
(27) Gao, J.-M.; Yang, S.-X.; Qin, J.-C. Azaphilones: Chemistry and Biology. *Chem. Rev.* **2013**, *113*, 4755-4811.
(28) Thines, E.; Anke, H.; Sterner, O. Trichoflectin, a Bioactive Azaphilone from the Ascomycete Trichopezizella nidulus. *J. Nat. Prod.* **1998**, *61*, 306-308.
(29) Zheng, Y.; Zhang, Y.; Chen, D.; Chen, H.; Lin, L.; Zheng, C.; Guo, Y. Monascus Pigment Rubropunctatin: A Potential Dual Agent for Cancer Chemotherapy and Phototherapy. *J. Agric. Food. Chem.* **2016**, *64*, 2541-8.
(30) Wang, W.; Liao, Y.; Chen, R.; Hou, Y.; Ke, W.; Zhang, B.; Gao, M.; Shao, Z.; Chen, J.; Li, F. Chlorinated Azaphilone Pigments with Antimicrobial and Cytotoxic Activities Isolated from the Deep Sea Derived Fungus Chaetomium sp. NA-S01-R1. *Mar. Drugs* **2018**, *16*.
(31) Park, J.-H.; Choi, G. J.; Jang, K. S.; Lim, H. K.; Kim, H. T.; Cho, K. Y.; Kim, J.-C. Antifungal activity against plant pathogenic fungi of chaetoviridins isolated from Chaetomium globosum. *FEMS Microbiol. Lett.* **2005**, *252*, 309-313.
(32) Winter, J. M.; Sato, M.; Sugimoto, S.; Chiou, G.; Garg, N. K.; Tang, Y.; Watanabe, K. Identification and Characterization of the Chaetoviridin and Chaetomugilin Gene Cluster in Chaetomium globosum Reveal Dual Functions of an Iterative Highly-Reducing Polyketide Synthase. *J. Am. Chem. Soc.* **2012**, *134*, 17900-17903.
(33) Makrerougras, M.; Coffinier, R.; Oger, S.; Chevalier, A.; Sabot, C.; Franck, X. Total Synthesis and Structural Revision of Chaetoviridins A. *Org. Lett.* **2017**, *19*, 4146-4149.
(34) Peixoto, P. A.; Boulangé, A.; Ball, M.; Naudin, B.; Alle, T.; Cosette, P.; Karuso, P.; Franck, X. Design and Synthesis of Epicocconone Analogues with Improved Fluorescence Properties. *J. Am. Chem. Soc.* **2014**, *136*, 15248-15256.
(35) Pyser, J. B.; Baker Dockrey, S. A.; Benítez, A. R.; Joyce, L. A.; Wiscons, R. A.; Smith, J. L.; Narayan, A. R. H. Stereodivergent, Chemoenzymatic Synthesis of Azaphilone Natural Products. *J. Am. Chem. Soc.* **2019**, *141*, 18551-18559.
(36) Hsu, W.-H.; Chen, T.-H.; Lee, B.-H.; Hsu, Y.-W.; Pan, T.-M. Monascin and ankaflavin act as natural AMPK activators with PPARα agonist activity to down-regulate nonalcoholic steatohepatitis in high-fat diet-fed C57BL/6 mice. *Food Chem. Toxicol.* **2014**, *64*, 94-103.
(37) Hsu, L.-C.; Liang, Y.-H.; Hsu, Y.-W.; Kuo, Y.-H.; Pan, T.-M. Anti-inflammatory Properties of Yellow and Orange Pigments from Monascus purpureus NTU 568. *J. Agric. Food Chem.* **2013**, *61*, 2796-2802.
(38) Kozakai, R.; Ono, T.; Hoshino, S.; Takahashi, H.; Katsuyama, Y.; Sugai, Y.; Ozaki, T.; Teramoto, K.; Teramoto, K.; Tanaka, K.; Abe, I.; Asamizu, S.; Onaka, H. Acyltransferase that catalyses the condensation of polyketide and peptide moieties of goadvionin hybrid lipopeptides. *Nat. Chem.* **2020**, *12*, 869-877.

(39) Xie, X.; Watanabe, K.; Wojcicki, W. A.; Wang, C. C. C.; Tang, Y. Biosynthesis of Lovastatin Analogs with a Broadly Specific Acyltransferase. *Chem. Biol.* **2006**, *13*, 1161-1169.
(40) Greene, N. P.; Crow, A.; Hughes, C.; Koronakis, V. Structure of a bacterial toxin-activating acyltransferase. *Proc. Natl. Acad. Sci.* **2015**, *112*, E3058-E3066.
(41) Musiol-Kroll, E. M.; Wohlleben, W. Acyltransferases as Tools for Polyketide Synthase Engineering. *Antibiot. (Basel)* **2018**, *7*, 62.
(42) Schmidt, N. G.; Pavkov-Keller, T.; Richter, N.; Wiltschi, B.; Gruber, K.; Kroutil, W. Biocatalytic Friedel–Crafts Acylation and Fries Reaction. *Angew. Chem. Int. Ed.* **2017**, *56*, 7615-7619.
(43) Baker Dockrey, S. A.; Lukowski, A. L.; Becker, M. R.; Narayan, A. R. H. Biocatalytic site- and enantioselective oxidative dearomatization of phenols. *Nat. Chem.* **2018**, *10*, 119-125.
(44) Zallot, R.; Oberg, N. O.; Gerlt, J. A., 'Democratized' genomic enzymology web tools for functional assignment. *Curr. Opin. Chem. Biol.* **2018**, *47*, 77-85.
(45) Gerlt, J. A., Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence–Function Space and Genome Context to Discover Novel Functions. *Biochemistry* **2017,** *56*, 4293-4308.
(46) Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L., Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta. Proteins Proteom.* **2015**, *1854*, 1019-1037.
(47) Jia, L.; Tu, X.; He, K.; Wang, C.; Yin, S.; Zhou, Y.; Chen, W. Monascorubrin and rubropunctatin: Preparation and reaction characteristics with amines. *Dyes Pigm.* **2019**, *170*, 107629.
(48) Huang, Z.; Zhang, S.; Xu, Y.; Li, L.; Li, Y. Structural characterization of two new orange pigments with strong yellow fluorescence. *Phytochem. Lett.* **2014**, *10*, 140-144.

Chapter 5: Conclusion and Future Directions

Advances in molecular biology and biochemistry within the last few decades have provided organic chemists with access to synthetically useful enzymes from myriad biosynthetic pathways.[1-18] Despite this, there is still a great deal to discover in the field of biocatalysis, requiring constant investigation of new enzymes and the reactions they catalyze. The body of work in this thesis seeks to provide access to molecular scaffolds with greater efficiency than attainable with analogous human-made methods using chemoenzymatic and multi-enzyme transformations. This Chapter summarizes the results of the preceding three chapters describing the identification and characterization of several known and unknown FDMOs, their application in the total synthesis of azaphilone natural products, and their combination with an AT enzyme to affect a one-pot sequence towards analogs of tricyclic azaphilones.

5.1 Conclusions

Specifically, we investigated enzymes from PKS pathways, beginning with a study on FDMOs performing dearomatization reactions. Dr. Attabey Rodríguez Benítez applied a state of the art bioinformatic analysis, called an SSN, to search and organize the enormous quantity of sequence data available online from the FDMO Pfam. Using this tool, she identified seven previously uncharacterized FDMOs with potential for synthetic utility. Together, Dr. Rodríguez Benítez, Dr. Summer Baker Dockrey, and I investigated

the stereo- and regioselectivity of these enzymes in a model hydroxylative dearomaitzation reaction. Ultimately, we demonstrated that the SSN is a useful tool in predicting the regioselectivity of these enzymes, but discovered that the stereoselectivity of this transformation is more intricately controlled than can be predicted with the SSN. Follow up work by Dr. Rodríguez Benítez supports our findings from this initial study.[19]

Following identification of the biocatalysts AzaH and AfoD, Dr. Baker Dockrey and I applied them to the chemoenzymatic total synthesis of three azaphilone natural products: trichoflectin, deflectin-1a, and lunatoic acid A. Careful analysis of the enantioenriched natural products from this study, in addition to calculated CD spectra and an X-ray crystal structure provided by Dr. Leo Joyce and Dr. Ren Wiscons, respectively, supports the need for structural revision of trichoflectin and deflectin-1a. These findings corrected an error in the azaphilone literature surrounding the original absolute configuration assignment of several azaphilone natural products.[20] Total synthesis and characterization of lunatoic acid A confirms its original absolute configuration was assigned correctly.

Finally, with a robust platform for oxidative dearomatization using FDMOs established, I sought to apply these enzymes to the total synthesis of other tricyclic azaphilones through a one-pot dual-enzyme sequence. To provide access to these more densely functionalized natural product scaffolds, I investigate the AT PigD for it's ability to mediate an acylation reaction of the dearomatized intermediate provide by AzaH. Once reaction conditions were optimized, I demonstrated PigD's promiscuity towards several dearomatized orcinaldehyde analogs, in addition to showing its compatibility with the AzaH reaction conditions. With assistance from Dr. Ye Wang, these two enzymes were

applied in sequence on a preparative scale, in combination with a thioester ACP mimic, to give access to the natural product rubropunctatin and several analogs thereof. NMR studies of these compounds suggests the acylated intermediate undergoes a subsequent cyclization reaction with unexpected regioselectivity, granting access to the linear tricyclic azaphilone scaffold with unprecedented selectivity.

5.2 Future Directions

Having developed novel strategies to access azaphilone natural product analogs, we aim to continue exploring the substrate scope and improving the yield of the aforementioned one-pot dual-enzyme platform. This will include screening the transformation against other aromatic substrates containing broader structural diversity than those investigated in Chapter 4, in addition to investigating PigD's ability to accept other acyl-donor reagents. In pursuit of greater atom economy, we seek to develop the use of truncated thioesters (such as *N*-acylcysteamine and thiophenol adducts) in place of the reported pantetheine-based thioester. We will also investigate acyl-donors with carbon chains of varying length and substitution to uncover the true substrate scope of PigD.

To access tricyclic natural product analogs containing the opposite C7 stereochemistry to those provided by AzaH, additional ATs will be identified through SSN analysis. We anticipate that investigation of enzymes related to PigD will uncover a protein capable of acylating dearomatized intermediates constructed by AfoD, ultimately providing increased control over the absolute configuration of the synthesized scaffolds. Once this route is established, it will be used to construct a large library of these tricyclic

azaphilones for SAR studies in collaboration with the Sexton Lab at the University of Michigan.

Regarding the question surrounding the actual substrate for PigD and the mechanism for the observed annulation regioselectivity, synthetic standards of the proposed acylated/uncyclized intermediates are currently being synthesized by Dr. Ye Wang for further investigation. We anticipate that experiments with these substrates, in addition to others possessing different functional groups and properties, will help uncover whether the observed selectivity is mediated by PigD or if the selectivity is conferred by the dearomatized substrate itself.

We envision that continued investigation of azaphilone compounds and the enzymes responsible for their construction will provide a means for additional development of these natural products towards potent pharmacophores. With newfound access to both enantioenriched azaphilone cores and the linear tricyclic azaphilone scaffold provided by the methods we've established, we aim to construct other related, highly decorated natural products and subsequent analogs using these techniques. We plan to modify our azaphilone tricycles through chemoenzymatic sequences to further diversify our growing library of compounds for biological investigation.

Towards this goal, we speculate that other enzymes from the rubropunctatin biosynthetic pathway could be of particular interest, given their role in the biosynthesis of related natural products.[21] For example: the oxidoreductases MrPigE and MrPigF, and the putative dehydrogenase MrPigH, are reported to manipulate the π system of the azaphilone core through reduction or oxidation of specific C–C bonds selectively, a feat that would be challenging with conventional chemical reagents.[21] Given the observed

compatibility of the enzymes AzaH and PigD, we hypothesize that additional enzymes like these could be added to our current dual-enzyme system, allowing for the electronics of the azaphilone core to be tuned and for new analogs to be generated in one-pot. Examining and comparing tricyclic analogs with varying degrees of oxidation decorating the basic core in SAR studies could also help identify how these structural changes affect the biological activity of these molecules. This modular, biomimetic synthetic platform could provide streamlined access to novel chemical space in the form of new natural product analogs from this underexplored class, aiding in our search for medicinal small molecules.

5.2 References

(1) Kelly, S. A.; Mix, S.; Moody, T. S.; Gilmore, B. F. Transaminases for industrial biocatalysis: novel enzyme discovery. *Appl. Microbiol. Biotechnol.* **2020,** *104* 4781-4794.

(2) Fisher, B. F.; Snodgrass, H. M.; Jones, K. A.; Andorfer, M. C.; Lewis, J. C. Site-Selective C-H Halogenation Using Flavin-Dependent Halogenases Identified via Family-Wide Activity Profiling. *ACS Cent. Sci.* **2019**, *5*, 1844-1856.

(3) Siloto, R. M. P.; Weselake, R. J. Site saturation mutagenesis: Methods and applications in protein engineering. *Biocatal. Agric. Biotechnol.* **2012**, *1*, 181-189.

(4) Winkler, C. K.; Schrittwieser, J. H.; Kroutil, W. Power of Biocatalysis for Organic Synthesis. *ACS Cent. Sci.* **2021**.

(5) Heckmann, C. M.; Paradisi, F. Looking Back: A Short History of the Discovery of Enzymes and How They Became Powerful Chemical Tools. *Chem. Cat. Chem.* **2020**, *12*, 6082-6102.

(6) Hughes, G.; Lewis, J. C. Introduction: Biocatalysis in Industry. *Chem. Rev.* **2018**, *118*, 1-3.

(7) Wandrey, C.; Liese, A.; Kihumbu, D. Industrial Biocatalysis: Past, Present, and Future. *Org. Process Res. Dev.* **2000**, *4*, 286-290.

(8) Bornscheuer, U. T.; Buchholz, K. Highlights in Biocatalysis – Historical Landmarks and Current Trends. *Eng.Life Sci.* **2005**, *5*, 309-323.

(9) Rodríguez Benítez, A.; Narayan, A. R. H. Frontiers in Biocatalysis: Profiling Function across Sequence Space. *ACS Cent. Sci.* **2019,** *5*, 1747-1749.

(10) Devine, P. N.; Howard, R. M.; Kumar, R.; Thompson, M. P.; Truppo, M. D.; Turner, N. J. Extending the application of biocatalysis to meet the challenges of drug development. *Nat. Rev. Chem.* **2018**, *2*, 409-421.
(11) Expanding biocatalysis for a sustainable future. *Nat. Catal.* **2020**, *3*, 179-180.
(12) Altreuter, D. H.; Clark, D. S. Combinatorial biocatalysis: taking the lead from Nature. *Currn. Opin. Biotechnol.* **1999**, *10*, 130–136.
(13) Rich, J. O.; Michels, P. C.; Khmelnitsky, Y. L. Combinatorial biocatalysis. *Curr. Opin. Chem. Biol.* **2002**, *6*, 161–167.
(14) Liu, W.; Wang, P. Cofactor regeneration for sustainable enzymatic biosynthesis. *Biotechnol. Adv.* **2007**, *25*, 369-384.
(15) Reetz, M. T. Biocatalysis in Organic Chemistry and Biotechnology: Past, Present, and Future. *J. Am. Chem. Soc.* **2013**, *135*, 12480-12496.
(16) Hughes, D. L. Biocatalysis in Drug Development—Highlights of the Recent Patent Literature. *Org. Process Res. Dev.* **2018**, *22*, 1063-1080.
(17) Akoh, C. C.; Chang, S.-W.; Lee, G.-C.; Shaw, J.-F. Biocatalysis for the Production of Industrial Products and Functional Foods from Rice and Other Agricultural Produce. *J. Agric. Food Chem.* **2008**, *56*, 10445-10451.
(18) Abdelraheem, E. M. M.; Busch, H.; Hanefeld, U.; Tonin, F. Biocatalysis explained: from pharmaceutical to bulk chemical production. *React. Chem. Eng.* **2019**, *4*, 1878-1894.
(19) Rodríguez Benítez, A.; Tweedy, S. E.; Baker Dockrey, S. A.; Lukowski, A. L.; Wymore, T.; Khare, D.; Brooks, C. L.; Palfey, B. A.; Smith, J. L.; Narayan, A. R. H. Structural Basis for Selectivity in Flavin-Dependent Monooxygenase-Catalyzed Oxidative Dearomatization. *ACS Catal.* **2019**, *9*, 3633-3640.
(20) Whalley, W. B.; Ferguson, G.; Marsh, W. C.; Restivo, R. J., The chemistry of fungi. Part LXVIII. The absolute configuration of (+)-sclerotiorin and of the azaphilones. *J. Chem. Soc., Perkin Trans. 1* **1976**, *13*, 1366-1369.
(21) Chen, R.; Liu, Q.; He, Y.; He, K.; Ding, X.; Kang, L.; Guo, X.; Xie, N.; Zhou, Y.; Lu, Y.; Cox, R. J.; Molnár, I.; Li, M.; Shao, Y.; Chen, F. Orange, red, yellow: biosynthesis of azaphilone pigments in Monascus fungi. *Chem. Sci.* **2017**, *8*, 4917-4925.